

Neuronal representation and attentional modulation of space and feature information in primate vision

Dissertation

to acquire the doctoral degree in mathematics and natural science

‘Doctor of Philosophy (Ph.D.)’

at the Georg-August-Universität Göttingen

in the doctoral degree programme Göttingen Graduate School for

Neurosciences, Biophysics, and Molecular Biosciences (GGNB)

at the Georg-August University School of Science (GAUSS)

Submitted by

Cheng Xue

from Wuxi, China

completed in Göttingen, September 2016

in final form in October 2017

Thesis Committee:

Prof. Dr. Stefan Treue (supervisor),

Cognitive Neuroscience Laboratory, German Primate Center

Prof. Dr. Fred Wolf,

Theoretical Neurophysics Group, Max Planck Institute for Dynamics and Self-Organization

Prof. Dr. Melanie Wilke,

Institute for Cognitive Neurology, University Medical Center Göttingen

Members of the examination board:

Referee: **Prof. Dr. Stefan Treue,**

Cognitive Neuroscience Laboratory, German Primate Center

Co-referee: **Prof. Dr. Fred Wolf,**

Theoretical Neurophysics Group, Max Planck Institute for Dynamics and Self-Organization

Other members of the Examination Board:

Prof. Dr. Melanie Wilke,

Institute for Cognitive Neurology, University Medical Center Göttingen

Dr. Igor Kagan,

Cognitive Neuroscience Laboratory, German Primate Center

Prof. Dr. Tim Gollisch,

Department of Ophthalmology, University Medical Center Göttingen

Prof. Dr. Ralf Heinrich,

Department of Cellular Neurobiology, Schwann-Schleiden Research Centre

Date of the oral examination: **19.10.2016**

Herewith I declare that I have written this thesis independently and with no other aids and sources than quoted

Göttingen, 03.09.2016

A handwritten signature in black ink, appearing to read 'Cheng Xue' in a cursive style.

(Cheng Xue)

Acknowledgements

I could not have completed this thesis without the support I am so lucky to have in various aspects. My most sincere gratitude goes to the following wonderful individuals.

For the past five years, my supervisor Prof. Dr. Stefan Treue has been a role model for me, both as a scientist and as a person. I appreciate the good research atmosphere in the lab he maintains as a group leader. His scientific guidance is everywhere in and beyond this thesis. The way of critical thinking I learned from him will be an invaluable fortune.

I am glad to have Antonino Calapai as my closest working partner in the lab. As a team, we have gone through many things together, and always supported each other in science and in life. I am also lucky to work with Dr. Suresh Krishna, who, as an important collaborator and an office mate, has been an always available source of advices (scientific and beyond) within five meters of distance.

Besides, there are many other colleagues who directly helped my thesis work. Dr. Philipp Schwedhelm helped building the experimental setup with stereoscopic display. Dr. Cliodhna Quigley worked with me on the spike waveform analysis. Dr. Sonia Baloni recorded the dataset upon which my work in chapter 3 is based. Kristin Dannhäuser and Julius Krumbiegel helped collect part of the data shown in chapter 4. I also appreciate Dr. Niklas Wilming and others for giving me the opportunity to collaborate on an important topic presented in chapter 5. I also enjoyed discussions with Dr. Moein Esghaei on neuronal synchronization, which led to our joint commentary in chapter 6. My thesis committee member Melanie Wilke and Fred Wolf participated in the constructive discussions in my thesis committee meeting. Finally, thanks to Vera Veith, Tao Yao and Benedict Wild for proof-reading parts of the thesis.

I should also thank our professional technician team: Ralf Brockhausen for software and hardware support of experimental setups; Dirk Prüsse, Sina Plümer, Leonore Burchardt for animal handling; Klaus Heisig for mechanical engineering; Beatrix Glaser for administrative affairs.

Last but not least, a special thanks to my family. My parents have always been encouraging and supporting me ever since I was born. Since 2012, my wife Qian Xue has embraced all my short-comings and decided to spend her life with me. I am blessed with such a powerful and capable woman, who, especially for the last couple of months before this thesis is completed, has taken charge of virtually everything at home, cared for the children while also finishing her own studies. I also owe a lot to my daughters Zhenzhen and Jingjing. You will have your dad back soon.

Content

<u>CHAPTER</u>	<u>PAGE</u>
1. General introduction	1
1.1 Primate visual system	2
1.2 Visual attention	7
1.3 Neuronal burst	12
1.4 Microsaccade.....	14
2. Motion and disparity in macaque area MST are independent from one another.....	27
3. Spatial attention reduces burstiness in macaque visual cortical area MST.....	63
4. Sustained spatial attention accounts for the direction bias of human microsaccades.....	75
5. Differential contribution of low- and high-level image content to eye movements in monkeys and humans.....	100
6. Does correlated firing underlie attention deployment in frontal cortex?.....	115
7. Summary and discussion.....	118
Curriculum vitae	121

Authors' contribution

Motion and disparity in macaque area MST are independent from one another

Antonino Calapai (AC), Stefan Treue (ST) and Cheng Xue (CX) designed the experiment; AC and CX performed the experiment and analysed the data; AC implemented the reverse correlation analysis underlying figure 1-5; CX implemented the population decoding analysis underlying figure 6-7; AC, CX and ST interpreted the data and wrote the paper.

Spatial attention reduces burstiness in macaque visual cortical area MST

Sonia Baloni Ray (SBR), Daniel Kaping (DK), and ST designed the experiment; SBR and DK performed the experiment; CX and B. Suresh Krishna designed the analysis; CX and BSK analyzed the data; CX, BSK and ST wrote the paper.

Sustained spatial attention accounts for the direction bias of human microsaccades

CX, AC, and ST designed the experiment; CX and Julius Krumbiegel (JK) performed the experiment; CX analyzed the data; CX, AC, and ST wrote the paper.

Differential contribution of low- and high-level image content to eye movements in monkeys and humans

Niklas Wilming (NW), Tim Kietzman (TK), and Peter Koenig (PK) designed the experiment; NW, Megan Jutras (MJ), and CX performed the experiment; NW and TK analyzed the data; All authors wrote the paper.

Does correlated firing underlie attention deployment in frontal cortex?

Moein Esghaei (ME) and CX contributed equally to the writing of the opinion piece

Chapter 1

General introduction

Visual perception lays the groundwork for many activities of animals and human. This ranges from the most basic skills of predator detection and preying to more advanced applications in the civilization world, such as driving, reading, etc., It also contributes to the foundation for various other higher cognitive functions such as decision-making and classification. Physiologically, perception entails not only the imagery representation of the outer world in the brain, but also attentional modulation of the representation, which is a vital cognitive process to selectively process visual inputs that are potentially relevant.

The past three decades has witnessed a huge expansion of our knowledge about vision. The first chapter of the thesis will be dedicated to give a structured summary of the findings and theories for the biological mechanism of visual perception. Next, this thesis endeavors to further extend our knowledge about visual perception in several aspects. In the second chapter, I present a study using electrophysiological recording in non-human primates to explore how multiple feature dimensions (such as location, moving direction and speed of a moving object) can together be encoded in a single visual area of the brain. While space and feature seem to be represented in the sensory system in a similar fashion, in the third chapter, again based on monkey

electrophysiological recordings, I demonstrate that the attentional modulation of spatial information recruits a qualitatively different mechanism than the attentional modulation of feature information. Mechanistically, one property that may differentiate spatial attention and feature-based attention is that the former is closely tied with the eye movement system. The fourth chapter presents a study in one type of small eye movements of human subjects during attempted gaze fixation (microsaccades). The results show a reliable correlation between the directions of microsaccades and attended location, with potential confounds excluded. As studies with monkey electrophysiology and human psychophysics are converging it is becoming an increasingly pressing agenda to evaluate the similarities of the visual selection mechanisms between humans and monkeys, and thereby to understand to which extent monkey electrophysiological findings can be translated into human vision. In the fifth chapter, I present the largest comparison to date between visual selection processes in humans and monkeys. With analysis on gaze positions during free viewing of images and computational modeling, we found that despite the vastly different behavioral repertoires of the two species, their visual selection behavior is mostly very similar with very limited differences in the contribution of presumably high-level image features. The sixth chapter contains an published article of my opinion on the role of cross-areal firing correlation on attention.

1.1 Primate visual system

The primate visual system is extraordinarily efficient in providing relevant information

in great detail when constantly faced with a massive amount of sensory input flow of around 400 megabytes per second (1). Various features of the object of interest, including location, shape, and direction and speed of its movement, can all be captured with a glimpse that takes no more than a fraction of a second. This is accomplished in two stages: Primary vision, in which the physical transformation from photons absorbed by sensors into electrical signals and transmitted to cerebral cortex; and higher vision, in which the visual information is decoded into feature representations, modulated by behavioral context, and integrated to form visual perception (2).

1.1.1 Retina and receptor distribution

The photoreceptors of our eyes are located in the retina, which is in the innermost part of the eye. The density of the photoreceptors is, however, not evenly assigned to every portion of our visual scene. The fovea, which refers to the central region of the retina that expands approximately 1.2 mm in diameter, has the highest density of photoreceptors. Therefore, the foveal visual space is represented with much higher visual acuity four times as high as the visual space merely 6 degrees visual angle away from the fovea (3). Because of this restricted visual acuity in fovea, we constantly move our eyes to bring objects of interest into our gaze so that their images fall on the fovea. The dynamics of foveation of visual stimuli in a visual scene would therefore reveal much about the visual selection mechanisms. Chapter 4 shows an example of such studies, in which by comparing the foveation behavior of humans

and monkeys during free viewing of pictures, we get an insight into the similarity between the visual selection mechanisms of the two species genetically optimized to live in vastly different environments.

The topographic spatial relationships are represented in the relative anatomical positions of the photoreceptors in retina(4): anatomically nearby neurons have overlapping yet slightly different RFs. Such a topographical relationship between anatomical positions and visual RFs, also known as retinotopy, is preserved as visual information is passed on to the cortex.

1.1.2 Dorsal and ventral pathway

The visual information picked up by the retina goes through a relay of optic nerves, the subcortical lateral geniculate nucleus (LGN), and arrives at the primary visual cortex (V1) in the occipital cortex. From V1, the visual information is further passed on to a hierarchy of visual cortical areas, which integrate increasingly more information and show more advanced processing. A popular hypothesis suggests that these visual areas can be functionally divided into a ventral stream and a dorsal stream. The ventral visual processing stream starts from area V1 then to area V2, 4 and goes further into the inferior part of the temporal lobe. These areas are mainly responsible for the representation and recognition of objects (5); and are therefore also referred to as the 'what' pathway. The latter, the dorsal stream, also starts from area V1, V2, then on to MT and MST, and goes further into the parietal lobe. These

areas are thought to be responsible for analyzing motion and relative spatial position of objects (6); and are therefore referred to as the 'where' pathway. This hypothesis of functional division is supported by the anatomical connections between these visual areas, differences in electrophysiological response properties, and the effects of cortical lesions(7).

For the physiology part of this thesis (chapter 2 and 3), I focus my study on area MT and MST in the dorsal stream, as example areas that may lead to general insights about the neuronal representation and selection mechanism for visual information.

1.1.3 Area MT / MST and motion perception

The medial temporal area (MT) and the medial superior temporal area (MST) are important loci in the dorsal 'where' pathway, which encode motion-related information. Area MT lies on the lower banks of the superior temporal sulcus (STS) of a macaque brain. It receives strong projections from V1(8). Compared to V1, MT is also retinotopically organized, but with receptive fields (RF) about ten times as large as those of V1(5). Like some V1 neurons, the majority of neurons in MT are found to be tuned to linear motion directions, i.e. for a certain MT neuron, only stimuli moving at a certain range of directions would induce its peak response(9, 10). Besides, some MT neurons are also tuned to other dimensions of motion-related feature, such as motion speed(11, 12), and the distance between the moving object and the observers' eyes (binocular disparity, (13, 14)). The integration of these information is necessary for the

perception of motion. Indeed, monkey electrophysiology studies found that the reliability and sensibility of motion direction discrimination decoded from activities of MT neurons are similar to those of the subjects *per se* (15). Lesion studies also confirm that impairment in macaque area MT compromises the subjects' capability of detection or discrimination of motion stimuli (16). Furthermore, by directly electro-stimulating the cluster of MT neurons with similar motion-direction preference, the subjects' perception of motion direction is biased towards the preferred direction of the stimulated neurons (17, 18). These evidences speak strongly in favor of MT's crucial role in our visual perception of motion.

Area MST lies on the anterior bank of STS. Neurons in MST typically have considerably larger RF compared to those of MT neurons at the same eccentricity (19-21). MST neurons are also found to encode the features represented in area MT (linear motion direction (22); linear motion speed (23); binocular disparity, (24)). Furthermore, neurons especially on the dorsal side of MST also encode more complicated stimuli, such as optic flow (22). Interestingly, one study also found that some MST neurons have opposite motion direction preference in near and far space with respect to the gaze position, and thereby might play a role in the subject's perception of self-motion with respect to the object in gaze (24). Yet, such a disparity dependent tuning of linear motion directions not found in area MT (14), and it is not clear how much this subset of individual MST cells actually contribute to the population representation of the features in area MST. In chapter 2, I present a

systematic study in the neuronal population representation of motion direction, motion speed and stimulus disparity in area MST, and interactions among the encoding of these features. We find that motion direction and disparity are the two most dominating features in explaining the variance of MST population activity, and the representations of these two features are mostly independent. The representation of motion speed, however, is dependent on disparity, i.e. the speed encoded in MST depends not only on how fast the stimuli sweep across the retina, but also on how far they are away from the viewer, arguing for an advanced role of MST in motion speed perception.

1.2 Visual attention

The ability to effectively process the visual information about our surrounding environment is important for species ranging from humans to insects(25, 26).Our visual system is equipped with hard-wired mechanisms that enhance the processing of salient or behaviorally relevant visual input and withdraw processing resources from the remaining inputs(27, 28).For instance, visual attention is found to improve the processing of task-relevant spatial locations and visual features (such as a particular motion direction or color) that leads to improved visual performance at these spatial locations and features(29-32). Meanwhile, associated with the perceptual improvements induced by spatial and feature-based attention, electrophysiology studies on non-human primates also identified a range of attentional effects on the neuronal firing rate(29, 33-35), the temporal structure of spike trains(36), the

correlated activities between neurons(37, 38), and the local field potential (LFP) (39, 40). These effects have been hypothesized to improve the sensory representation of attended stimuli by enhancing neural responses and by reducing noises among neurons that represent the attended locations and/or features.

1.2.1 Spatial attention vs. feature-based attention

As mentioned above, attention can selectively enhance sensory processing of different aspects of the visual world. This selective mechanism can apply to various dimensions. For instance, spatial attention enables selective processing of specific locations against others(33, 41, 42); while feature-based attention controls selection of a certain stimulus feature such as color and orientation (28, 43).A unifying hypothesis, called the feature-similarity gain model, proposes that spatial and feature-based attention act via similar neural mechanisms even if they may affect different brain areas and connections(34, 44-47).In other words, in the context of visual attention, space can be regarded as just another visual feature.

Existing neural data have been supporting or generally consistent with this model. Specifically, attention enhances the firing rate of MT neurons representing attended spatial locations in comparison to those representing unattended locations (33);,similarly, the firing rate gain of MT neurons representing an attended visual feature is enhanced compared to that of neurons representing an unattended visual feature(28).Attention can also shift the tuning curve of firing rate towards the attended

spatial location (48) or feature (49). However, besides the firing rate modulations, attention has also been shown to produce a variety of other neural effects, which have been demonstrated for feature-based attention. For example, spatial attention has been shown to modulate both inter-neuronal correlation (50), spike temporal structure(36), and the oscillatory properties of spike trains(39). Spatial attention also affects the spectral content of LFP as well as the synchrony and phase-relationship between LFP and spikes. However, these neural effects for feature-based attention have not been tested so far.

In chapter 3, I investigated the effects of spatial and feature-based attention on the tendency of macaque area MST neurons to fire consecutive action potentials (burstiness).. Against the predictions from the unifying approach of the feature-similarity gain model, I found that spatial and feature-based attention evoke qualitatively different effects on the burstiness. Therefore, the results in chapter 3 indicate that the feature-similarity gain model has only restricted applicability, attention to spatial and feature information may not necessarily employ the same neuronal mechanism. In that sense, space might not just be another feature for all aspects of attentional modulation.

1.2.2 Overt shift of spatial attention

The location of gaze determines which parts of our visual environment are processed with high-accuracy foveal vision. As we explore the visual scene, objects that attract

our attention are sequentially brought into our gaze by means of fast eye movements(or saccades). Since gaze positions can only shift from one location to another in a serial fashion, it is critical for a biological organism to effectively prioritize different locations in a cluttered visual scene based on their behavioral relevance (e.g. to detect prey or predators). Therefore, studying this underlying cortical selection process could potentially reveal, on a behavioral level, how humans selectively direct attention(51).

Over the years, researchers have been exploring the overt visual selection process from two distinct approaches. On one hand, to investigate 'what' is selected,behavior and eye movement dynamics have been measured, mostly from human subjects (52-54). These studies typically correlated the scan paths of the gaze with the features in the visual scene presented to the subjects, so as to understand the characteristics of objects that attract overt attention in general, or in a certain behavioral context. The abundance of data in this research direction yields successful mathematic models that predict the sequence of overt spatial attention selection (55, 56). On the other hand, to investigate 'how' the selection is biophysically implemented, electrophysiology studies directly measured the neuronal activities during the visual selection process (35, 57), performed in macaque monkeys, the most prominent model system for studying human cognition. Studies in this direction have correlated attentional selection with modulations in various aspects (57-59).

However, before combining the two lines of studies to infer the neuronal mechanism of visual attention selection, there is just one important gap to fill: how can we assume such mechanisms are in humans and monkeys? Despite the genetic similarity between the two species, the behavioral repertoires are vastly different between human and non-human primates in many respects. It is therefore highly likely that eye movement behaviors and visual selection dynamics are also different between the two species. Yet, besides several proofs of concept (60-62), the extent of the differences and their potential impact on translating data from monkeys to humans are not systematically evaluated.

In chapter 4, I present a collaborative study, in which we performed the largest comparison to date between the scan paths of humans and rhesus monkeys when presented with standard image sets of urban scenes, natural scenes, and fractal images. Predictions based on computational models of visual saliency trained by human and monkey data indicate that the attentional selection process in both species are largely determined by low-level selection mechanisms, with only a small contribution by higher-level selection mechanisms, among which differences exist between human and monkey.

1.2.3 The exogenous vs. endogenous spatial attention

There are two mechanisms through which spatial attention is deployed, initially proposed in light of findings from experiments (63). The “exogenous” attention is an

automatic process that shifts rapidly towards salient stimuli or events (64, 65). This mechanism is largely an involuntary reflex. On the other hand, the “endogenous” attention is a volitional, goal-directed process that allocate more cognitive resource to the behavioral relevant location, independent of visual stimuli (66). Although both mechanisms induce similar attentional enhancement in behavior and neuronal activity, the temporal dynamics of engagement is different. Their distinct time evolution patterns have been demonstrated in behavioral data (67), and neuronal activity in visual cortex (68). These studies show that exogenous attention acts faster than endogenous attention, but is easily disrupted by behavioral irrelevant stimuli. Alongside with the fact that various neural disorders affect the two mechanisms differently (69, 70), it seems that different networks are at play for the two modes of attention.

1.3 Neuronal Bursts

Extensive electrophysiology studies, both *in vitro* and *in vivo*, have reported the tendency of neurons to fire brief periods of spikes in quick succession (burstiness)(71, 72). Burstiness is believed to be associated with a variety of physiological processes, such as synapse formation (73) and long-term potentiation (74). Analysis of bursting activity has also been used as an important tool in applications such as studying the impact of genetic or chemical manipulations on network activity (75, 76).

Recently, it has been shown in V4, a key locus in the ventral stream of visual cortical

information processing, that attention can also modulate aspects of neuronal firing patterns that operate on a fast timescale: burstiness, defined as the tendency of a neuron to discharge consecutive spikes at very short inter-spike intervals, decreases in the broad-spiking neurons of area V4 when spatial attention is directed into their receptive fields (RF)(36). Though the specific functional consequence of this attentional modulation remains unknown, the effect is intriguing, because the functional properties and neural utility of bursts in spike-trains has been a topic of much speculation and interest (77-79). A current and plausible hypothesis states that bursts enhance information transfer because neuronal inputs composed of closely spaced spikes are more efficient at driving post-synaptic neurons which act as coincidence detectors because of their short integration time-constants (74). As pointed out by Anderson et al. (36), this hypothesis predicts that to drive downstream neurons more efficiently, burstiness would increase when attention is directed towards a neuron's RF. However, the burstiness reduction observed indicates the opposite.

At present, it remains unclear if the effect of spatial attention on burstiness is restricted to the ventral pathway or even only V4 and whether it extends to other types of attention. Further, though it has been recently proposed based on a computational model that the effects of spatial attention on burstiness and firing rate emerge from a common mechanism (36), there is no empirical data on how the attentional modulation of burstiness relates to the well-known modulation of firing rate by attention. To address this, we analyzed extracellular single-neuron recordings from

area MST of two rhesus monkeys performing a spatial and feature-based attention task. Both shifting spatial attention into the RF and deploying feature-based attention to the preferred direction (relative to the non-preferred direction) enhanced the firing rate of MST neurons, as expected based on previous studies (28, 33, 80). In addition, spatial attention also led to a concurrent net reduction in burstiness, as reported earlier from V4. However, feature-based attention did not modulate burstiness, though it did enhance firing-rate. This lack of effect on burstiness for feature-based attention is not explained by its smaller effects on firing-rate (compared to spatial attention). Further, the effects of spatial attention on firing rate and burstiness could be dissociated. Our results extend our understanding of the attentional effects on the temporal patterns of action potential discharge and support the idea that different types of attention may involve different physiological mechanisms.

1.4 Microsaccade

As is described in a previous section, we move our eyes to bring important visual stimuli under gaze, where the sensory resolution is the highest. Besides these voluntary eye movements, however, our eyes are still in constant motion. Even when we are intently maintaining our gaze at a certain location, miniature, involuntary eye movements never stop. According to their magnitudes and velocity profiles, These fixational eye movements can basically be classified into three types (81). The type of fixational eye movement with smallest overall amplitude is tremor (or nystagmus), which refers to the constant, rhythmical oscillation ($\sim 90\text{Hz}$) that occur to each of the

eyes independently (82). Simultaneously occurring with tremor, another much slower type of fixational eye movement is drift, which slowly shifts the fixated image across a dozen photoreceptors on the retina, due to the instability of oculomotor system (83). Drifts are occasionally interrupted by microsaccades, the third type of fixational eye movements, which are abrupt binocular eye movements that share most of the properties with saccades except that they are involuntary and relatively smaller in magnitude (84).

Although fixational eye movements can partially be attributed to noise in the oculomotor system, microsaccades in particular are often suggested, although not without controversies, to be functionally relevant. For instance, microsaccades correct accumulated drifts away from the gaze location (85) and counteract perceptual fading due to neural adaptation (86). However, it has been pointed out, that microsaccades are not necessary, either for maintaining fixation or to keep the world visible (87).

The more compelling behavioral relevance of microsaccades came from several independent studies that revealed a correlation between microsaccades and shifts in attention (88-91). Specifically, these studies reported a consistent bias of microsaccade direction, immediately following a spatial cue. However, this effect alone is not enough to conclude microsaccade directions as an index for spatial attention, without ruling out two major confounds, concerning the internal and the external cause of microsaccade direction effect. First, covert visual spatial attention is

often entangled with saccade planning (92, 93). With evidences also pointing towards a common biological mechanism for saccades and microsaccades(94), it seems more plausible that microsaccade is a direct correlate of saccade planning. Without a systematic study designed to disentangle attention from saccade planning, it remains elusive which one internally drives the microsaccade direction effect. Second,the microsaccade direction effect can, alternatively, be directly driven by the external cue, rather than by attention (91). Considering that all the studies that reported the microsaccade direction effect focus specifically on the time period immediately after the cue (~250-300ms), and that exogenous and endogenous cues have opposite effects on microsaccade direction (91, 95), it is in doubt whether sustained internal attention alone, without the presence of external cues, can generate this effect.

In Chapter 4, I endeavor to address these two major concerns with two psychophysics experiments with human subjects. I found that sustained internal attention by itself, not entangled with saccade planning and without the presence of any external cue, can bias microsaccade direction towards the attended location. This finding echoes with previous findings and strongly argues for microsaccades as a reliable index of endogenous attention.

References:

1. Schacter D. Psychology. Second Edition ed: Worth Publishers; 2011.
2. Bear MF. Neuroscience: Exploring the Brain. Philadelphia, PA: Lippincott Williams

& Wilkins; 2007.

3. Purves D, Augustine GJ, Fitzpatrick D, Hall WC, LaMantia AS, McNamara JO, et al. Neuroscience. 3rd ed. ed. Sunderland, Massachusetts U.S.A.: Sinauer Associates; 2004.
4. Wandell BA, Brewer AA, Dougherty RF. Visual field map clusters in human cortex. *Philos Trans R Soc Lond B Biol Sci.* 2005;360(1456):693-707.
5. Mishkin M, Ungerleider LG. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav Brain Res.* 1982;6(1):57-77.
6. Van Essen DC, Maunsell JH, Bixby JL. The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization. *J Comp Neurol.* 1981;199(3):293-326.
7. Goodale MA, Milner AD. Separate visual pathways for perception and action. *Trends Neurosci.* 1992;15(1):20-5.
8. Ungerleider LG, Mishkin M. The striate projection zone in the superior temporal sulcus of *Macaca mulatta*: location and topographic organization. *J Comp Neurol.* 1979;188(3):347-66.
9. Dubner R, Zeki SM. Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Res.* 1971;35(2):528-32.
10. Albright TD. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol.* 1984;52(6):1106-30.

11. Maunsell JH, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol.* 1983;49(5):1127-47.
12. Liu J, Newsome WT. Functional organization of speed tuned neurons in visual area MT. *J Neurophysiol.* 2003;89(1):246-56.
13. Maunsell JH, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol.* 1983;49(5):1148-67.
14. Smolyanskaya A, Ruff DA, Born RT. Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. *J Neurophysiol.* 2013;110(12):2806-16.
15. Newsome WT, Britten KH, Movshon JA. Neuronal correlates of a perceptual decision. *Nature.* 1989;341(6237):52-4.
16. Newsome WT, Pare EB. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J Neurosci.* 1988;8(6):2201-11.
17. Salzman CD, Britten KH, Newsome WT. Cortical microstimulation influences perceptual judgements of motion direction. *Nature.* 1990;346(6280):174-7.
18. Salzman CD, Murasugi CM, Britten KH, Newsome WT. Microstimulation in visual area MT: effects on direction discrimination performance. *J Neurosci.* 1992;12(6):2331-55.
19. Komatsu H, Wurtz RH. Relation of cortical areas MT and MST to pursuit eye movements. III. Interaction with full-field visual stimulation. *J Neurophysiol.*

1988;60(2):621-44.

20. Komatsu H, Wurtz RH. Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *J Neurophysiol.* 1988;60(2):580-603.

21. Born RT, Tootell RB. Segregation of global and local motion processing in primate middle temporal visual area. *Nature.* 1992;357(6378):497-9.

22. Graziano MS, Andersen RA, Snowden RJ. Tuning of MST neurons to spiral motions. *J Neurosci.* 1994;14(1):54-67.

23. Wurtz RH, Yamasaki DS, Duffy CJ, Roy JP. Functional specialization for visual motion processing in primate cerebral cortex. *Cold Spring Harb Symp Quant Biol.* 1990;55:717-27.

24. Roy JP, Komatsu H, Wurtz RH. Disparity sensitivity of neurons in monkey extrastriate area MST. *J Neurosci.* 1992;12(7):2478-92.

25. Carrasco M. Visual attention: the past 25 years. *Vision Res.* 2011;51(13):1484-525.

26. Wiederman SD, O'Carroll DC. Selective attention in an insect visual neuron. *Curr Biol.* 2013;23(2):156-61.

27. Reynolds JH, Chelazzi L. Attentional modulation of visual processing. *Annu Rev Neurosci.* 2004;27:611-47.

28. Treue S, Martinez Trujillo JC. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature.* 1999;399(6736):575-9.

29. Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annu*

Rev Neurosci. 1995;18:193-222.

30. Treue S. Neural correlates of attention in primate visual cortex. Trends Neurosci. 2001;24(5):295-300.

31. Moore T, Armstrong KM. Selective gating of visual signals by microstimulation of frontal cortex. Nature. 2003;421(6921):370-3.

32. Bichot NP, Heard MT, DeGennaro EM, Desimone R. A Source for Feature-Based Attention in the Prefrontal Cortex. Neuron. 2015;88(4):832-44.

33. Treue S, Maunsell JH. Attentional modulation of visual motion processing in cortical areas MT and MST. Nature. 1996;382(6591):539-41.

34. Treue S, Maunsell JH. Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. J Neurosci. 1999;19(17):7591-602.

35. Bisley JW. The neural basis of visual attention. J Physiol. 2011;589(Pt 1):49-57.

36. Anderson EB, Mitchell JF, Reynolds JH. Attention-dependent reductions in burstiness and action-potential height in macaque area V4. Nat Neurosci. 2013;16(8):1125-31.

37. Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. Nat Neurosci. 2011;14(7):811-9.

38. Oemisch M, Westendorff S, Everling S, Womelsdorf T. Interareal Spike-Train Correlations of Anterior Cingulate and Dorsal Prefrontal Cortex during Attention Shifts. J Neurosci. 2015;35(38):13076-89.

39. Fries P. Neuronal gamma-band synchronization as a fundamental process in

cortical computation. *Annu Rev Neurosci.* 2009;32:209-24.

40. Esghaei M, Daliri MR, Treue S. Attention Decreases Phase-Amplitude Coupling, Enhancing Stimulus Discriminability in Cortical Area MT. *Front Neural Circuits.* 2015;9:82.

41. McAdams CJ, Maunsell JH. Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron.* 1999;23(4):765-73.

42. Reynolds JH, Chelazzi L, Desimone R. Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci.* 1999;19(5):1736-53.

43. Motter BC. Neural correlates of feature selective memory and pop-out in extrastriate area V4. *J Neurosci.* 1994;14(4):2190-9.

44. Maunsell JH, Treue S. Feature-based attention in visual cortex. *Trends Neurosci.* 2006;29(6):317-22.

45. Martinez-Trujillo JC, Treue S. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol.* 2004;14(9):744-51.

46. Maljkovic V, Nakayama K. Priming of pop-out: II. The role of position. *Percept Psychophys.* 1996;58(7):977-91.

47. Bundesen C. A theory of visual attention. *Psychol Rev.* 1990;97(4):523-47.

48. Womelsdorf T, Anton-Erxleben K, Pieper F, Treue S. Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat Neurosci.* 2006;9(9):1156-60.

49. David SV, Hayden BY, Mazer JA, Gallant JL. Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron.* 2008;59(3):509-21.

50. Cohen MR, Maunsell JH. Using neuronal populations to study the mechanisms underlying spatial and feature attention. *Neuron*. 2011;70(6):1192-204.
51. Petersen SE, Posner MI. The attention system of the human brain: 20 years after. *Annu Rev Neurosci*. 2012;35:73-89.
52. Treisman AM, Gelade G. A feature-integration theory of attention. *Cogn Psychol*. 1980;12(1):97-136.
53. Braun J, Julesz B. Withdrawing attention at little or no cost: detection and discrimination tasks. *Percept Psychophys*. 1998;60(1):1-23.
54. Betz T, Kietzmann TC, Wilming N, König P. Investigating task-dependent top-down effects on overt visual attention. *J Vis*. 2010;10(3):15 1-4.
55. Itti L, Koch C. Computational modelling of visual attention. *Nat Rev Neurosci*. 2001;2(3):194-203.
56. Wilming N, Betz T, Kietzmann TC, König P. Measures and limits of models of fixation selection. *PLoS One*. 2011;6(9):e24038.
57. Zhou H, Desimone R. Feature-based attention in the frontal eye field and area V4 during visual search. *Neuron*. 2011;70(6):1205-17.
58. Buschman TJ, Miller EK. Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron*. 2009;63(3):386-96.
59. Eimer M. The neural basis of attentional control in visual search. *Trends Cogn Sci*. 2014;18(10):526-35.
60. Shepherd SV, Steckenfinger SA, Hasson U, Ghazanfar AA. Human-monkey gaze

correlations reveal convergent and divergent patterns of movie viewing. *Curr Biol.* 2010;20(7):649-56.

61. Einhauser W, Kruse W, Hoffmann KP, König P. Differences of monkey and human overt attention under natural conditions. *Vision Res.* 2006;46(8-9):1194-209.

62. Berg DJ, Boehnke SE, Marino RA, Munoz DP, Itti L. Free viewing of dynamic stimuli by humans and monkeys. *J Vis.* 2009;9(5):19 1-5.

63. Jonides J. In: Long J, Baddeley A, editors. *Attention and Performance*. Hillsdale, NJ: Erlbaum; 1981. p. 187-204.

64. Irwin DE, Colcombe AM, Kramer AF, Hahn S. Attentional and oculomotor capture by onset, luminance and color singletons. *Vision Res.* 2000;40(10-12):1443-58.

65. Yantis S, Jonides J. Abrupt visual onsets and selective attention: evidence from visual search. *J Exp Psychol Hum Percept Perform.* 1984;10(5):601-21.

66. Posner MI. *Chronometric Explorations of Mind* Hillsdale, NJ: Erlbaum; 1978.

67. Müller HJ, Rabbitt PM. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *J Exp Psychol Hum Percept Perform.* 1989;15(2):315-30.

68. Busse L, Katzner S, Treue S. Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. *Proc Natl Acad Sci U S A.* 2008;105(42):16380-5.

69. Renner P, Grofer Klinger L, Klinger MR. Exogenous and endogenous attention orienting in autism spectrum disorders. *Child Neuropsychol.* 2006;12(4-5):361-82.

70. Danckert J, Maruff P, Crowe S, Currie J. Inhibitory processes in covert orienting in

- patients with Alzheimer's disease. *Neuropsychology*. 1998;12(2):225-41.
71. Weyand TG, Boudreaux M, Guido W. Burst and tonic response modes in thalamic neurons during sleep and wakefulness. *J Neurophysiol*. 2001;85(3):1107-18.
 72. Pasquale V, Martinoia S, Chiappalone M. A self-adapting approach for the detection of bursts and network bursts in neuronal cultures. *J Comput Neurosci*. 2010;29(1-2):213-29.
 73. Maeda E, Robinson HP, Kawana A. The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons. *J Neurosci*. 1995;15(10):6834-45.
 74. Lisman JE. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci*. 1997;20(1):38-43.
 75. Eisenman LN, Emnett CM, Mohan J, Zorumski CF, Mennerick S. Quantification of bursting and synchrony in cultured hippocampal neurons. *J Neurophysiol*. 2015;114(2):1059-71.
 76. Charlesworth P, Morton A, Eglén SJ, Komiyama NH, Grant SG. Canalization of genetic and pharmacological perturbations in developing primary neuronal activity patterns. *Neuropharmacology*. 2016;100:47-55.
 77. Bair W, Koch C, Newsome W, Britten K. Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J Neurosci*. 1994;14(5 Pt 1):2870-92.
 78. Krahe R, Gabbiani F. Burst firing in sensory systems. *Nat Rev Neurosci*. 2004;5(1):13-23.
 79. Izhikevich EM. *Dynamical Systems in Neuroscience*. London: The MIT Press;

2007.

80. Patzwahl DR, Treue S. Combining spatial and feature-based attention within the receptive field of MT neurons. *Vision Res.* 2009;49(10):6.

81. Martinez-Conde S, Macknik SL, Hubel DH. The role of fixational eye movements in visual perception. *Nat Rev Neurosci.* 2004;5(3):229-40.

82. Riggs LA, Ratliff F, Cornsweet JC, Cornsweet TN. The disappearance of steadily fixated visual test objects. *J Opt Soc Am.* 1953;43(6):495-501.

83. Carpenter RHS. *Movements of the eyes.* University of Michigan: Pion; 1988.

84. Rolfs M. Microsaccades: small steps on a long way. *Vision Res.* 2009;49(20):2415-41.

85. Engbert R, Kliegl R. Microsaccades keep the eyes' balance during fixation. *Psychol Sci.* 2004;15(6):431-6.

86. Martinez-Conde S, Macknik SL, Troncoso XG, Dyar TA. Microsaccades counteract visual fading during fixation. *Neuron.* 2006;49(2):297-305.

87. Collewijn H, Kowler E. The significance of microsaccades for vision and oculomotor control. *J Vis.* 2008;8(14):20 1-1.

88. Rolfs M, Laubrock J, Kliegl R. Shortening and prolongation of saccade latencies following microsaccades. *Exp Brain Res.* 2006;169(3):369-76.

89. Engbert R, Kliegl R. Microsaccades uncover the orientation of covert attention. *Vision Res.* 2003;43(9):1035-45.

90. Hafed ZM, Clark JJ. Microsaccades as an overt measure of covert attention shifts. *Vision Res.* 2002;42(22):2533-45.

91. Laubrock J, Engbert R, Kliegl R. Microsaccade dynamics during covert attention. *Vision Res.* 2005;45(6):721-30.
92. Rizzolatti G, Riggio L, Dascola I, Umiltà C. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia.* 1987;25(1A):31-40.
93. Smith DT, Schenk T. The Premotor theory of attention: time to move on? *Neuropsychologia.* 2012;50(6):1104-14.
94. Hafed ZM, Krauzlis RJ. Similarity of superior colliculus involvement in microsaccade and saccade generation. *J Neurophysiol.* 2012;107(7):1904-16.
95. Rolfs M, Engbert R, Kliegl R. Crossmodal coupling of oculomotor control and spatial attention in vision and audition. *Exp Brain Res.* 2005;166(3-4):427-39.

Motion and disparity in Macaque area MST are independent from one another

Antonino Calapai^{1*}, Cheng Xue^{1*}, Stefan Treue^{1,2}

¹ Cognitive Neuroscience Laboratory, German Primate Centre, Goettingen, Germany

² Faculty of Biology and Psychology, Goettingen University, Goettingen, Germany

* These authors contributed equally to this work.

Abstract

Within the visual cortex, information from sensory stimulation is first decomposed into features, represented by neurons in specialized visual areas, and later integrated to form a global percept. It has been suggested that at the processing level of macaque visual cortical area MST, the integration of the direction and the perceived distance of a moving stimulus, occurs; with such integration providing the basic computational input to the network responsible for self-motion perception. While the theory is elegant, the evidence for this process is rather scarce, with only few studies available in literature. Here, we recorded from area MST of gaze fixated awake macaque monkeys, while displaying stereoscopic random dot patch stimuli with various combinations of features. Surprisingly, we found that the interaction of motion direction and disparity did not explain more variance in the neuronal activity. In addition, on the population level, the decoding of motion direction seems to be rather independent from the decoding of disparity, suggesting that the integration of the two domains here considered, as basis for the computation of self-motion, is unlikely to take place in area MST.

Introduction

Amongst the over 30 visual processing areas identified in the macaque's cerebral cortex (Felleman & Van Essen, 1991), extrastriate areas V2, V3, V4, MT and MST (Brodmann areas 18 and 19) can be partitioned into two distinct pathways: the *form-colour* pathway (Zeki, 1978b; 1978a) and the *visual-motion* pathway (Maunsell & Van Essen, 1983c). Both pathways are traditionally thought to follow a serial and hierarchical functional organization, according to which, lower areas serve as computational node to the processing of higher areas, with a certain degree of reciprocity (Felleman & Van Essen, 1991), for a review see Perry & Fallah, 2014. While most of the areas comprising these two pathways seem well defined regarding their respective hierarchical function; along the visual-motion pathway, the medial superior temporal area (MST) shows rather diversified selectivity. In macaque monkeys, MST can be anatomically partitioned into two subareas with distinct functions: a dorsal portion (MSTd), mainly composed of neurons with large receptive fields and selectivity to the basic motion components of optic flow (expansion, contraction, rotation and translation); and a ventral portion (MSTl), composed of neurons with smaller receptive fields and selectivity to linear motion direction, much resembling the properties of MT neurons (Tanaka, Sugita, Moriya, & Saito, 1993). Given its complex architecture and functionality, human and macaque studies suggest MSTd's involvement in a number of processes: heading perception (Britten & van Wezel, 2002); integration of motion information through feature decomposition of optic flow (Duffy & Wurtz, 1991; Graziano, Andersen, & Snowden, 1994; Orban et al., 1992; Saito et al., 1986; Tanaka & Saito, 1989); inertial motion in darkness (Takahashi et al., 2007); perceptual cue integration (Gu, Angelaki, & DeAngelis, 2008); gaze stabilization in smooth pursuit (Kawano, Inoue, Takemura,

Kodaka, & Miles, 1999; Takemura, Inoue, Kawano, Quaia, & Miles, 2001); integration of vestibular and visual cues (Sakata, Shibutani, & Kawano, 1983); visual spatial attention (Treue & Maunsell, 1996); visual working memory (Mendoza-Halliday, Torres, & Martinez-Trujillo, 2014) and integration of colour (Perry & Fallah, 2014; Tchernikov & Fallah, 2010). Moreover, within the most studied domain - the sensitivity to visual motion - MST's neurons located in both anatomical subdivisions MSTl and MSTd encode multiple feature dimensions at once: motion directions in both the spiral space (Graziano et al., 1994; Mineault, Khawaja, & Butts, 2012) and the linear space (Saito et al., 1986); binocular disparities (Roy, Komatsu, & Wurtz, 1992; Takemura et al., 2001; Yang, Liu, Chowdhury, DeAngelis, & Angelaki, 2011); the speed of a given motion pattern (Maunsell & Van Essen, 1983a; Price & Born, 2013). While these tuning preferences are most often considered in isolation, the potential dependence of the encoding of one feature on another is still under considerable debate, and yet may reveal important functions.

Disparity-dependent direction selectivity

MST's sensitivity to binocular disparity - the difference between the right and left retinal projections of an object - has often been an influential factor in this area's motion selectivity, as well as vestibular selectivity. A currently leading hypothesis is that binocular disparity sensitivity and motion selectivity are functionally integrated at the processing level of MST to infer self-motion (Roy et al., 1992; Takemura et al., 2001; Yang et al., 2011). Cells showing direction-dependent disparity tuning (or DDD) in which the tuning for motion depends on the disparity value considered, have been reported in area MST. (Roy et al., 1992; Roy & Wurtz, 1990). However, the

reported proportions of DDD cells in this area vary considerably. Roy et al. observed DDD tuning in around 40% of MST cells, while Yang et al. reported it in around 5% of the cells analysed. Considering also that multiple studies have suggested the DDD cells do not exist in MT (DeAngelis & Newsome, 1999; Maunsell & Van Essen, 1983b; Smolyanskaya, Ruff, & Born, 2013) - an area in close functional and anatomical proximity to MST – it seems that DDD cells might be exclusive to MST.

The present study aims at shedding some light onto the functional relationship between disparity selectivity and motion directionality in macaque area MST, by focussing on two experimental questions. First, to characterize the area contribution in the estimation of self-motion, we determine the proportion of cells showing DDD tuning. Secondly we quantify the involvement of each feature dimension, as well as their joint contribution, in explaining the overall population response to ultimately address the role of area MST in the processing of these two features along the visual-motion pathway.

Materials and Methods

Single unit activity was recorded from two rhesus monkeys (*Macaca mulatta*, both male; monkey I 10-year-old, weighed 9 kg; monkey N, 16-year-old, weighed 10kg), implanted with custom made titanium headpost and recording chamber (19 mm diameter), over the superior temporal sulcus (monkey I on the left hemisphere, monkey N on the right hemisphere). Surgeries were performed under general anaesthesia and post-surgical care using standard techniques. All procedures were conducted in accordance with German laws governing animal care and approved by the district government of Oldenburg, Lower Saxony, Germany.

Setup

The animals were seated in a primate chair for the duration of the experimental session. The animals were positioned in front of a rear projection screen (dlp Black Bead, Denmark, 171.5 x 107.2 cm) so that the screen laid 104 cm from the animal's eyes. Stereoscopic visual stimulation was achieved by mean of two coupled projectors (Projection Design F22, Norway, 60 Hz refresh rate, 1920 x 1200 pixels) and circular polarization filters (SX42 – HD). Binocular crosstalk, as assessed by a spectroradiometer (SpectraScan PR-650, Photo Research, USA), was below the minimum measurable luminance of 0.2 foot-lambert (or 0.68 candela/meter²). Eye position was monitored with a binocular eye tracking system (Eyelink 1000, SR-Research, Canada) throughout the course of the experimental session at a sampling rate of 500 Hz.

Behavioral Tasks

Every recording session was comprised of two consecutive behavioural protocols. In the first part, we place a single probe stimulus at various locations to identify the neuron's receptive field (RF). Subsequently, in the second part, we characterized the neuron's response to visual stimuli placed at the centre of the RF, with various combination of motion and disparities. Basic behavioural requirements to the animals in the two protocols were identical: a red dot (2x2 degrees of visual angle - dva) placed at the centre of the projected screen, instructed the animal to engage eye fixation, and initiate the trial (monkey I by depressing a mechanical button, monkey N by touching a lever; both installed inside the primate chairs). The dim fixation point then lit up, signalling the animal that a new trial was about to start.

When, during the trial, the fixation point would dim down again, the animal was required to release the button, or turn the lever, within 500ms, to earn a drop of fluid reward. Breaking eye fixation at any time during a trial, reacting before a fixation dot dim, or fail to react to a fixation dot dim within the 500ms time window, would lead to the abortion of the trial and no reward would be delivered. Regardless of the outcome, after 1.5 seconds a new trial was presented. The mean reaction times were 290ms (sd 27 ms) for monkey I and 366ms (sd 25ms) for monkey N.

In the *mapping of the receptive field* protocol (RF protocol), upon correct initiation of the trial, a single random dot pattern (RDP, 4 dva in diameter, 20 dots, each measured 0.25 dva in diameter moving at speed of 10 dva/s, with zero-coherence in motion directions, at a luminance of 7.07 cd/m²) would appear for 3 frames (~50 ms) at a random position on the projection screen. The stimulus then disappeared and, after one blank frame (16.67 ms), reappeared at a different and randomized location. At a random point in time during RDPs flashing (between 1500 and 3500 ms from the appearance of the first stimulus), the dimming of the fixation point described above would occur. The behavioural protocol was terminated after reaching 150 successful trials, which resulted in 5850 probes presented, over an x and y space of 41 * 41 dva around the centre (0,0) of the horopter, with positive and negative values around the fixation position (x = from -10 to 30, y = from -20 to 20).

The characterization of the neuronal sensitivity to different visual features (Tuning Protocol), was carried out in direct succession of the receptive field mapping protocol. Upon receptive field identification, a single RDP (with full motion coherence, variable diameter adjusted to the receptive field size determined through

online analysis, 200 dots of 0.25 dva each, with an average luminance of 12.8 cd/m²), was placed at the centre of a neuron's RF and its x and y position was then kept constant throughout the experiment. The stimulus' motion domain (spiral or linear), motion direction (0, 45, 90, 135, 180, 235, 270, 315 degree, for linear motion, the values refer to the angles between dot velocity and the horizontal line; for spiral motion, the values refer to the angles between dot velocity and the radial line of the RDP aperture, see (Graziano et al., 1994)), binocular disparity (-2, -1.5, -1, -0.5, 0, 1, 1.5 degree) and speed (at 1 dva from RDP's centre), would rapidly and randomly change every 5 frames (83.33 ms). Here as in the RF protocol, the animal was required to depress the lever in within 500 ms after the dimming of the fixation point (between 1500 and 3500 ms from the appearance of the first stimulus). Each session of this experimental protocol requires 500 hit trials to complete, so that a total of ~13000 stimuli would be displayed. Considering the number of possible feature combinations (8 directions * 8 disparities * 8 speeds * 2 motion domains = 1024), each stimulus would be displayed for 12 repetitions on average.

Data Collection

The recording electrodes (platinum/tungsten cores, quartz insulated, Thomas Recording, Germany, and FHC, ME), single tip as well as four channels (impedance between 0.8 and 2.5 MΩ) were either loaded into a multi-electrode manipulator (Tetrode Mini Matrix System, Thomas Recording, Germany) or into a custom made guide tube held on a chamber grid. The respective recording device was mounted on the recording chamber of the animal, prior the recording session. Consequent to manual adjustment of the *medio lateral* and *anterior posterior* coordinates on the x-y

table of the manipulator, the guide tubes was manually lowered enough to penetrate the superficial tissue covering the dura. The micro-drive system of the manipulator, by mean of a dedicated motor controller, would then lower the electrodes at ~ 10 $\mu\text{m}/\text{second}$, upon regular impedance monitoring by the experimenter. Electrical signals were amplified and then recorded with a sampling rate of 40 kHz and 16-bit precision, using an Omniplex acquisition system (Plexon, USA). After recording, the raw signal acquired was filtered with a 6-pole Bessel high pass filter (250 Hz cut-off) using the OfflineSorter V3 software (Plexon, USA). Single units were identified as clusters of similar waveforms, crossing an individually set detection threshold, and separated from the main noise cluster in the space of the first two PCs (for a review see Lewicki, 1998). We thus isolated 229 cells for monkey I and 18 cells from monkey N, with 154 for monkey I and 10 for monkey N showing clear responses to visual stimulation.

Data Analysis

Both protocols, employing a rapid series of stimuli presentations, were optimized for *reverse correlation* analysis (Bair, Cavanaugh, Smith, & Movshon, 2002; Borghuis et al., 2003; Chichilnisky, 2001; de Boer & Kuyper, 1968; Ringach, Hawken, & Shapley, 1997), where any given spike train is probabilistically associated with individual stimulus features. Given a range of latencies, stretching from 300 milliseconds before the spike to 50 ms after the spike, binned in 5 ms steps, we implemented the reverse correlation by first counting the number of total occurrences of a certain stimulus category (for example expansion) at a given

latency relative to the spike and then dividing this sum by the total occurrences of all categories comprising the corresponding feature (for example spiral motion). For directionally selective cells, for example, this procedure outputs a probability value for each motion direction at each latency. Ultimately the results are interpreted as the likelihood of each feature category, at each latency considered, to have preceded each spike in the spike train. It is important to note that in such two dimensional space (latency vs category), the sum of the probability of all categories at any latency is always equal to 1.

Two-dimensional Gaussian for receptive field mapping

To quantitatively estimate the size and the distance of the receptive field from the fixation point, on a cell-by-cell basis, we first identified the latency yielding the highest variance of spike counts for all probe locations, and fit a 2 dimensional Gaussian of the following form:

$$G = B + \left(A * \exp \left(- \left(\frac{(x \cdot \cos\theta - y \cdot \sin\theta - x_0)^2}{2\sigma_x^2} + \frac{(y \cdot \cos\theta + x \cdot \sin\theta - y_0)^2}{2\sigma_y^2} \right) \right) \right)$$

where B is the baseline probability; A is the amplitude; x_0 and y_0 are the coordinate of the centre of the receptive field in degrees of visual angle; σ_x and σ_y are the standard deviation of the Gaussian in the two dimensions; θ is the orientation of the longer axis of the fitted ellipse. The size of the receptive field is defined as the area obtained considering 2 standard deviations and assuming an elliptical shape.

Piecewise Polynomial Interpolation for disparity tuning estimation

Disparity tuning of each cell was computed in MATLAB through a piecewise polynomial interpolation with a smoothing parameter of 0.99, using the built-in function *fitttype* under the mode 'SmoothingSpline'.

Von Mises fit for directionality estimation

The tuning of each neuron to the motion stimuli, for both the linear and the spiral domains, was computed by fitting the probabilities of each motion direction, derived by the reverse correlation of the each neurons' spike train, to a von Mises distribution, a circular approximation of the normal distribution (Berens, 2008; Mineault et al., 2012; Smolyanskaya et al., 2013; Takahashi et al., 2007), of the following form:

$$f(x | \mu, \kappa, a, b) = b + a * \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}$$

where μ and $1/\kappa$ represent preferred direction and variance, a and b amplitude and baseline probability and the component $I_0(\kappa)$ is the modified Bessel function of order 0.

Negative Binomial Regression Model

To assess the amount of variability explained by the motion and the disparity, on a cell by cell basis, we built four generalized additive models considering spike count as response variable and disparity, direction and their putative interaction, as

predictors. Model 1 assumes that motion direction does not contribute to the variance of spike count:

$$E_{sc} = \exp (\beta_0 + \beta_1 \cdot disparity)$$

Model 2 assumes that disparity does not contribute to the variance of spike count:

$$E_{sc} = \exp (\beta_0 + \beta_2 \cdot direction)$$

Model 3 assumes both disparity and direction contribute to the variance independently:

$$E_{sc} = \exp (\beta_0 + \beta_1 \cdot disparity + \beta_2 \cdot direction)$$

Model 4 further adds an interaction term between disparity and direction:

$$E_{sc} = \exp (\beta_0 + \beta_1 \cdot disparity + \beta_2 \cdot direction + \beta_3 \cdot disparity \cdot direction)$$

It is important to note that in models considering the contribution of motion direction (m2, m3 and m4), this circular covariate was linearized with a Von Mises transformation, by adding a squared covariate in the regression models. Note also that the spike count E_{sc} consisted of the total number of spikes occurring within an 80ms time window, shifted according to the latency yielding the highest variability (*optimal latency*) assessed through reverse correlation.

Principal component analysis for population decoding

In order to achieve the unsupervised clustering of feature domains analysis described in the results section, we first constructed a covariance matrix based on the spike count of the 154 cells we recorded from monkey I as variables, and the 512

stimuli of one stimulus category (linear or spiral) as observations. The covariance matrix is z-scored through observations, so as to normalize the neurons to their general firing rate. A principal component analysis (PCA) is then performed on the covariance matrix, using the build-in *pca* function of MATLAB. Once the clustering of stimuli in the subspace expanded by PCs were obtained, individual dots were marked post hoc according to stimuli features, so as to determine which stimulus feature drives the clustering. Finally, to quantify the performance of the classification between the clusters (as in Fig 3B), first the centroids of each category in the PC subspace were identified and then connected. Stimuli from the two categories were projected on this connecting axis and the area under receiver operative characteristic curve from the two distributions resulted in the performance of classification.

Results

General population statistics.

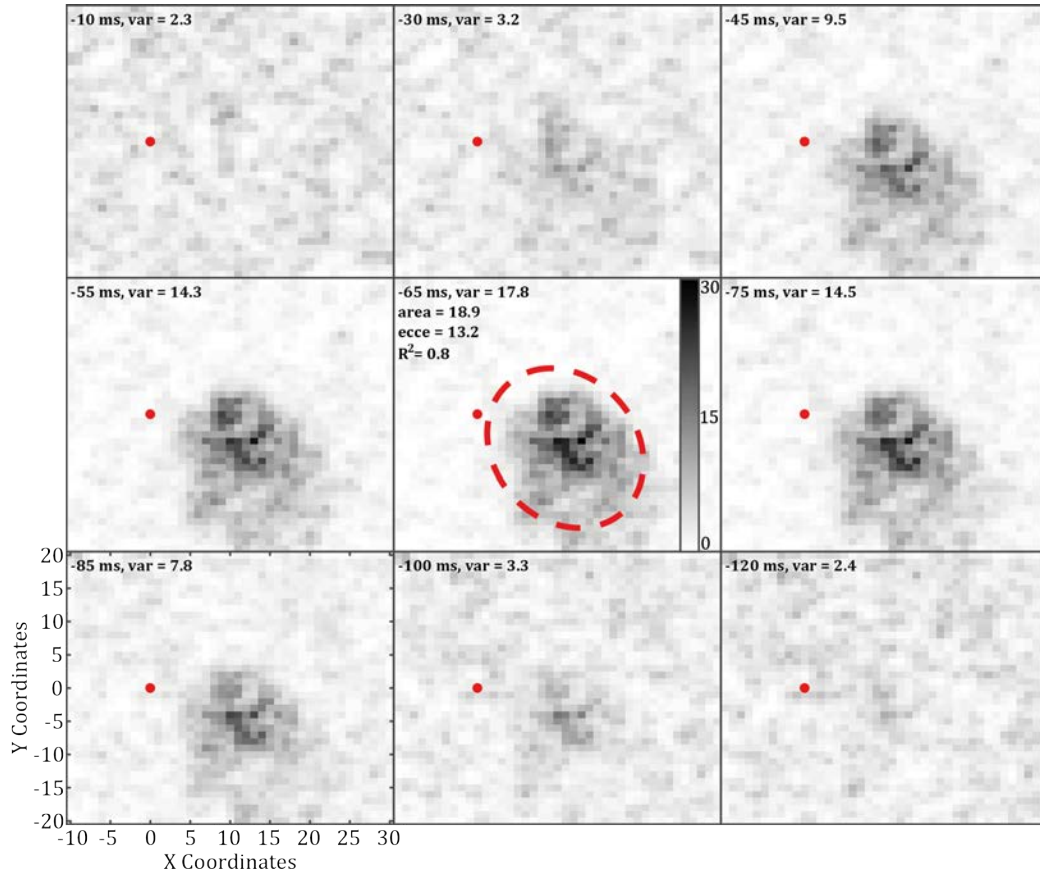


Figure 1 – Receptive field maps describing the dynamics of an example cell as assessed through reverse correlation and fit with a 2 dimensional Gaussian (see Methods – Data Analysis). Each subplot shows the spatial selectivity at incremental latencies. The greyscale map spans from white to black for low probability to high probability respectively. The array of probabilities depicted in the central plot, showing the latency containing the highest variance of the probabilities indicated at the top right of each panel (var), was fit with a 2 dimensional Gaussian to derive size (area) and eccentricity (ecce), in dva, with respect to the fixation point (red dot). Bar on central plot shows absolute count of occurrences of each location, from which probabilities are derived.

Of the 164 cells comprising the population in analysis, data to estimate the receptive field was available for 147 units. We applied a single inclusion criterion of an adjusted r-squared above 0.15, based on the fit of neuronal responses with a 2

dimensional Gaussian, to include only units for which at least 15% of the variance is explained. This reduced the data to dataset to 85 units, for which the size and location of the receptive field was computed (for monkey I receptive field population average is 20 dva, range 27 dva; average population eccentricity is dva, range 22 dva). Figure 1 illustrates the process of determining the receptive field dynamics for one example unit (cell-074-01+01-137.3), convoluted with a 3-by-3 kernel.

Throughout the 85 cells depicted in figure 2, no simple correlation was observed

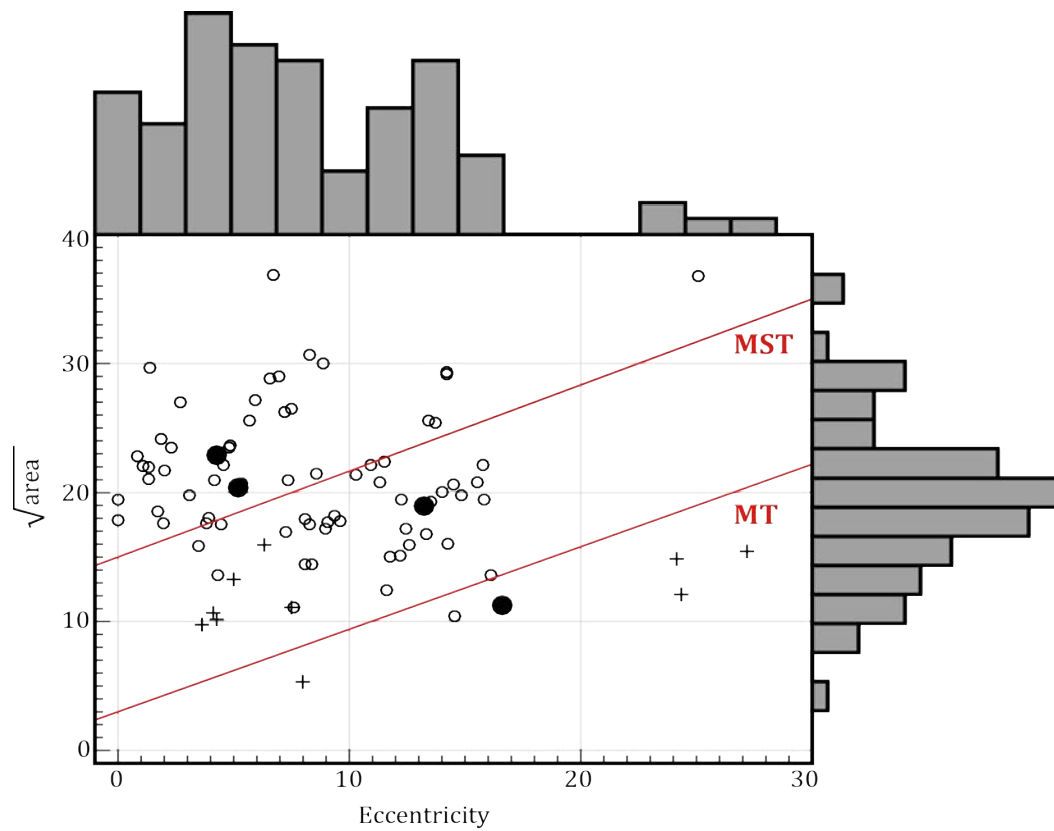


Figure 2 – Scatter plot and distribution histograms of receptive fields' size (square root of the area) and eccentricity for the 85 cells (75 from monkey I – circles, 10 from Monkey N – crosses) satisfying the inclusion criteria of adjusted $r^2 > 0.15$ to a 2 dimensional Gaussian fit. Red lines are derived from existing literature on MST and MT receptive field size and eccentricity (see results) and are here shown as reference for our data set. The filled circle indicates the example units.

between areas and eccentricities ($\rho = -0.06$, $p = 0.53$; Spearman's rank correlation test – all values are rounded to the next integer). While areas range from 10 to 37 dva, with an average value of 20 dva, eccentricities range from 0 to 22 dva, covering mostly the right hemifield, with several units coding for the foveal region and often crossing the midline, towards the ipsilateral visual field, as expected for MST neurons (Saito et al., 1986; Tanaka & Saito, 1989). In line with existing literature of anesthetized monkeys on single cell activity of area MST and MT (Desimone & Ungerleider, 1986), units described in this study show receptive fields' size and eccentricity spanning all the way from values almost approaching MT's typical ratio, at the low end of the spectrum, to MST's typical ratio and beyond (red lines in figure 2 are extracted from Desimone & Ungerleider, 1986 and represent best fitting regression lines for MT and MST, histologically identified).

Similarly to the example receptive field map shown in figure 1, figure 3 illustrates the process of characterizing motion and disparity selectivity for the same example unit (cell-074-01+01-137.3). For each given cell, upon identification of the latency yielding highest variance, a von Mises distribution was fit to the probability of each motion category for both motion domains (see Methods) to extract preferred direction. To ensure that only directional cells were included in the analysis, inclusion criteria was set to an adjusted r^2 above 0.64, through which were accepted 89 out of 164 cells for spiral motion and 115 out 164 for linear motion (some units satisfied the criteria only for one of the two motion domains). Figure 4 summarizes the distribution of the preferred directions, for the two motion domains. In line with previous literature (Duffy & Wurtz, 1991) there is no evidence for an underlying non-uniform distribution of the preferred directions as assessed by the Rayleigh statistical test for non-uniformity of circular data (for spiral $p = 0.059$, for linear $p =$

0.152). Suggesting that neurons were sampled from an area representing all the motion direction with the same likelihood.

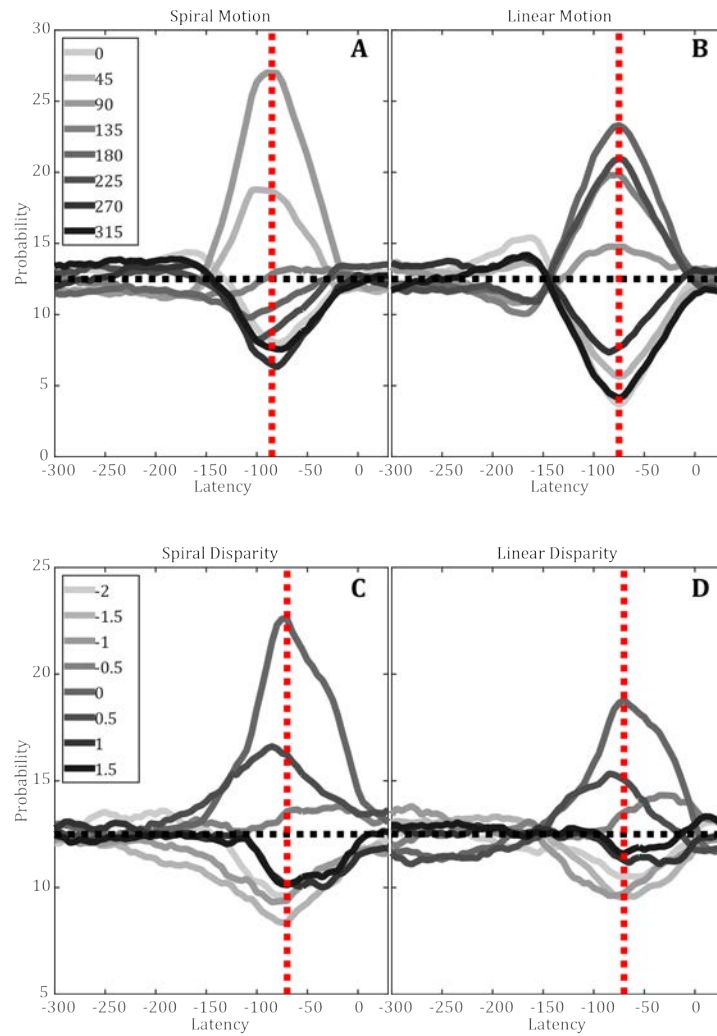


Figure 3 - Time course of motion directions and disparity selectivity for the example unit, based on 36194 spikes. Each subplot shows the probability of each motion category (A and B) or each disparity level (C and D) assessed in the spiral (A and C) and linear (B and D) domains, versus the temporal distance between each spike and each stimulus presentation. Latency 0 indicates simultaneous occurrence of spike and stimulus. Red dashed lines indicate the latency with the largest separation (highest variance), marking the time at which the unit shows optimal selectivity (A= -85 ms; B= -75 ms; C= -70 ms; D= -70 ms).

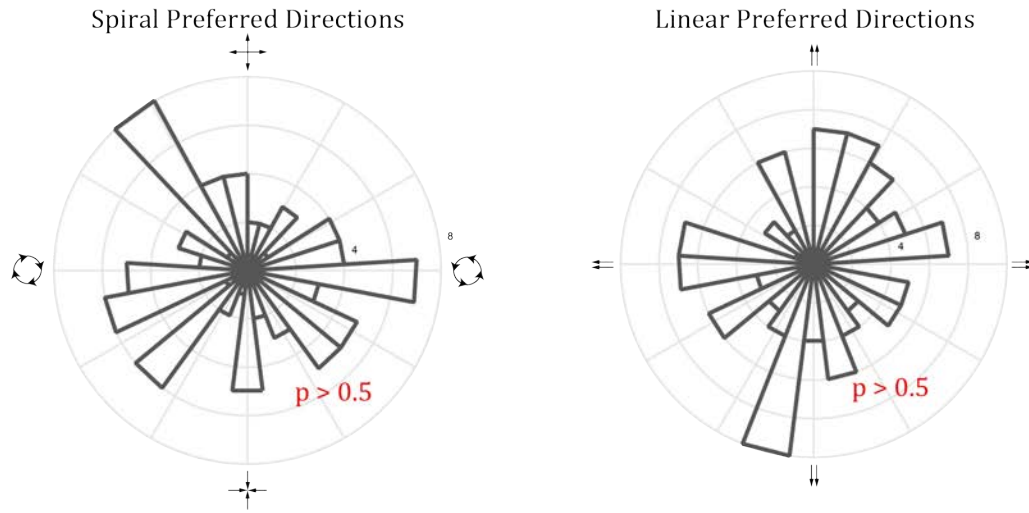


Figure 4 – Distribution of preferred directions for spiral motion (left) and linear motion (right). Only units with an adjusted r^2 above 0.64 are considered. P values, referring to Rayleigh test for non-uniformity of circular data.

Quantitative measurements of disparity selectivity throughout the population were based on two indices previously introduced by (Roy et al., 1992). Figure 5 shows the distribution of the *disparity index* (DI) derived with the formula:

$$DI = 1 - \left| \left(\frac{null - exp}{preferred - exp} \right) \right|$$

where *exp* represents the expected probability 0.125 (1 over 8, the number of disparities tested), *null* the lowest probability and *preferred* the highest. The resulting value indicates the strength of the disparity tuning for each given cell. Considering then the units with a disparity index above or equal to 0.2 (Roy et al., 1992), with the second index, the *zero index* (ZI), is possible to determine whether

the disparity selectivity refers to disparity zero (namely no binocular disparity) or either far or the near disparity, derived with the formula:

$$ZI = \frac{zero - exp}{max - exp}$$

where the term *zero* indicates the probability for the disparity value 0, *max* is the probability for any non-zero category and *exp* is again the expected probability of 0.125. As a result any cell yielding a ZI above 1 is considered tuned to 0 retinal disparity and conversely, any cell with a ZI equal or below 1 is considered to be tuned to either the far or the near space (Roy et al., 1992).

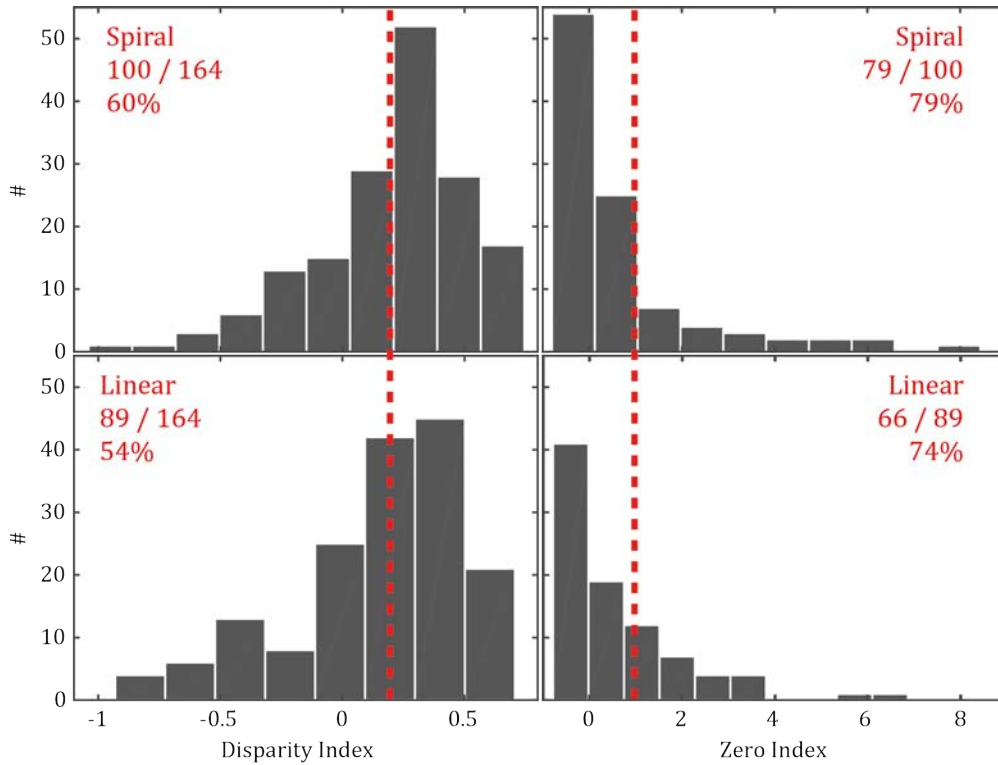


Figure 5 – Distributions of disparity index and zero index values across the population. Disparity index thresholds of 0.2 and zero index thresholds of 1 are indicated by the red dashed lines, while the ratios in red indicate the proportion of units falling respectively above and below the thresholds.

In line with previous reports (Maunsell & Van Essen, 1983a; 1983b; Takemura et al., 2001), but contrasting with other reports (Roy et al., 1992), 61% of the units recorded showed sensitivity to spiral disparity. 78% of these cells showed tuning to either far or near with equal proportion. 53% of the units showed sensitivity to linear disparity, with 74% tuned to either near or far space with equal proportion.

The population encode linear and spiral motion direction

In addition to single cell selectivity profiles, we also investigated: 1) which feature dimension can be decoded from the neuronal activities of the whole population; 2) if such decoders can be constructed for the features here considered; 3) how well can the decoders perform within each feature dimension and 4) how the different decoders relate to each other.

Specifically, we performed a principal component analysis (PCA) based on the spike counts of the 154 neurons recorded in monkey I in response to the 512 linear motion stimuli (see materials and methods). As a result, we obtain 154 principal components (PCs, weighted linear combinations of the 154 neurons), ranked by their contributions to the spike count variance across stimuli. Based on the responses of the first two PCs to the stimuli, we found that the 512 stimuli automatically formed eight clusters, which happened to align with the eight linear motion directions (Fig 6A). This made the combination of the first two PCs a very good decoder for motion direction. Based on the activities of these 154 neurons, it is also possible to reliably decode the direction of linear motion with the first two components (classification performances between neighbouring directions are all

above 93%, Fig 6B). Similarly, a separate PCA on the spiral motion stimuli also yielded a similar outcome. As a result motion directions can be decoded through an unsupervised clustering (Fig 6B).

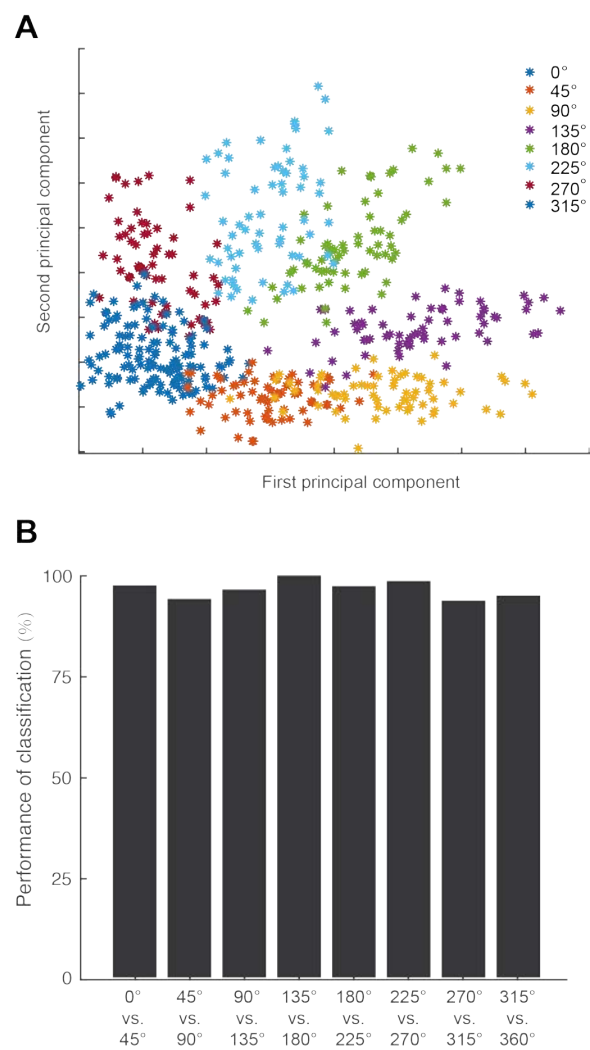


Figure 6 - The first two principal components decodes motion direction. A) the unsupervised clustering of stimuli in the subspace of the first two principal components. Each dot represents the neuronal response to a stimulus, projected on the first two principal components. The dots were coloured according to their spiral motion direction of the stimuli. B) the performance of classification between neighbouring motion directions.

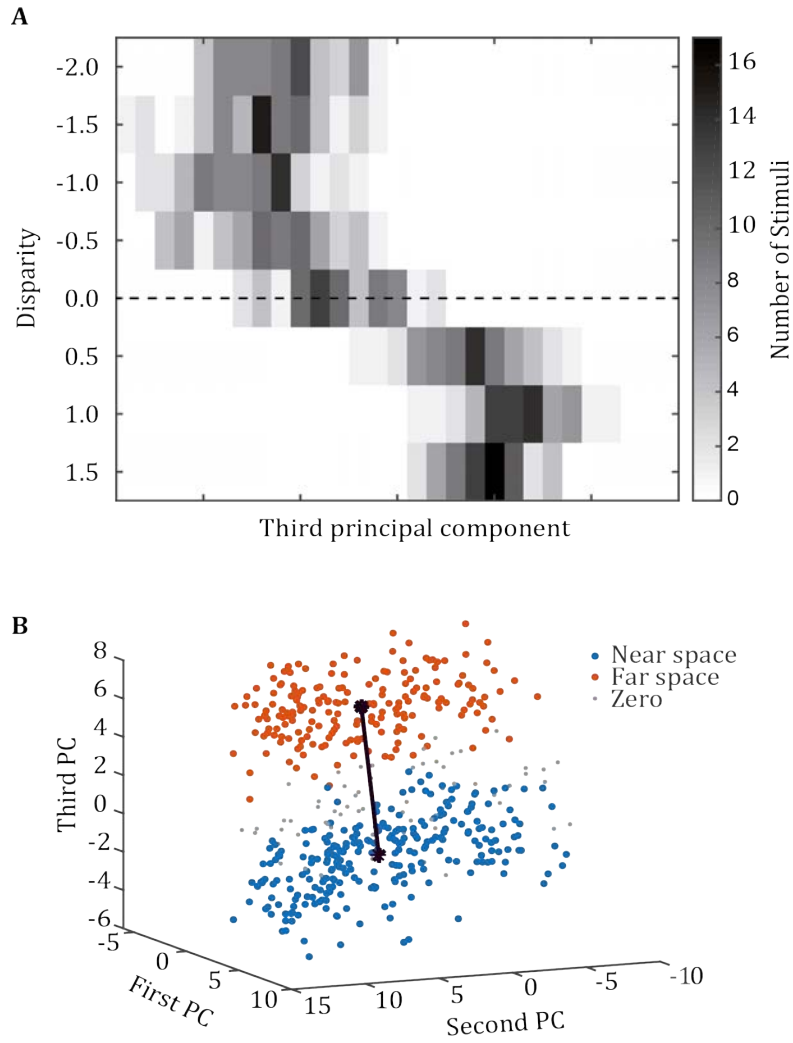


Figure 7 - The categorization of near and far stimuli is independent of the representation of motion direction. A) Near-far categorization with the third principal component. The greyscale map shows the distributions of stimuli with a certain disparity (vertical axis) on the third principal component (horizontal axis). B) The relationship between near-far decoder (third principal component) and motion direction decoder (first two principal components). Each dot represents the neuronal response to a stimulus, projected on the first three principal components. The dots were colored according to their relative depths with the fixation point. The difference between the center of near stimuli and far stimuli (black segment) is almost parallel to the third principal component axis.

The encoding of near-far categorization is independent from the encoding of other stimuli features

While the first two components encode the motion direction of the stimuli, the third component seems to be independently encoding disparity. Stimuli displayed in the near space (disparity smaller than zero) and far space (disparity bigger than zero) have distinct distributions in the third component (Fig 7A).

Furthermore, this representation of near-far categorization and the representation of motion direction are largely independent. As shown in Fig. 7B, In the 3-D space explained by the first three PCs, we obtained the centroids of dots representing near stimuli (blue) and dots representing far stimuli (red), and create a disparity axis connecting the two centroids (dark black line). The smaller the angle between a given PC and this axis, the larger the PC contribute to the near-far categorization. We found the disparity axis is almost perpendicular to the plane expanded by the first two PCs (88° , Fig. 7B), which contains the representation of motion direction; while the third PC alone contributed 99.8% to the disparity axis.

Such clear independence of disparity encoding from motion direction encoding solicits the question whether it is even possible to categorize near stimuli from far stimuli regardless of what type of motion is displayed (linear or spiral). To test this, we obtained the disparity decoder from a PCA with data from monkey I with linear motion stimuli, and applied it on the spike count data from the same animal but with spiral motion stimuli, so as to guess which of the spiral motion stimuli were displayed in the near space and which in the far space. Compared post hoc with the real disparity values, the guessed near-far categorization is 98.75% correct, which shows that the near-far categorization in area MST is largely agnostic about the motion domain of the stimulus.

Interdependence of features in area MST

In order to probe any putative dependence of motion selectivity upon disparity change, as previously investigated for MST and MT (Roy et al., 1992; Smolyanskaya et al., 2013), the joint probability of motion and disparity, independently for the two motion domains, was calculated at the best latency of the reverse correlation.

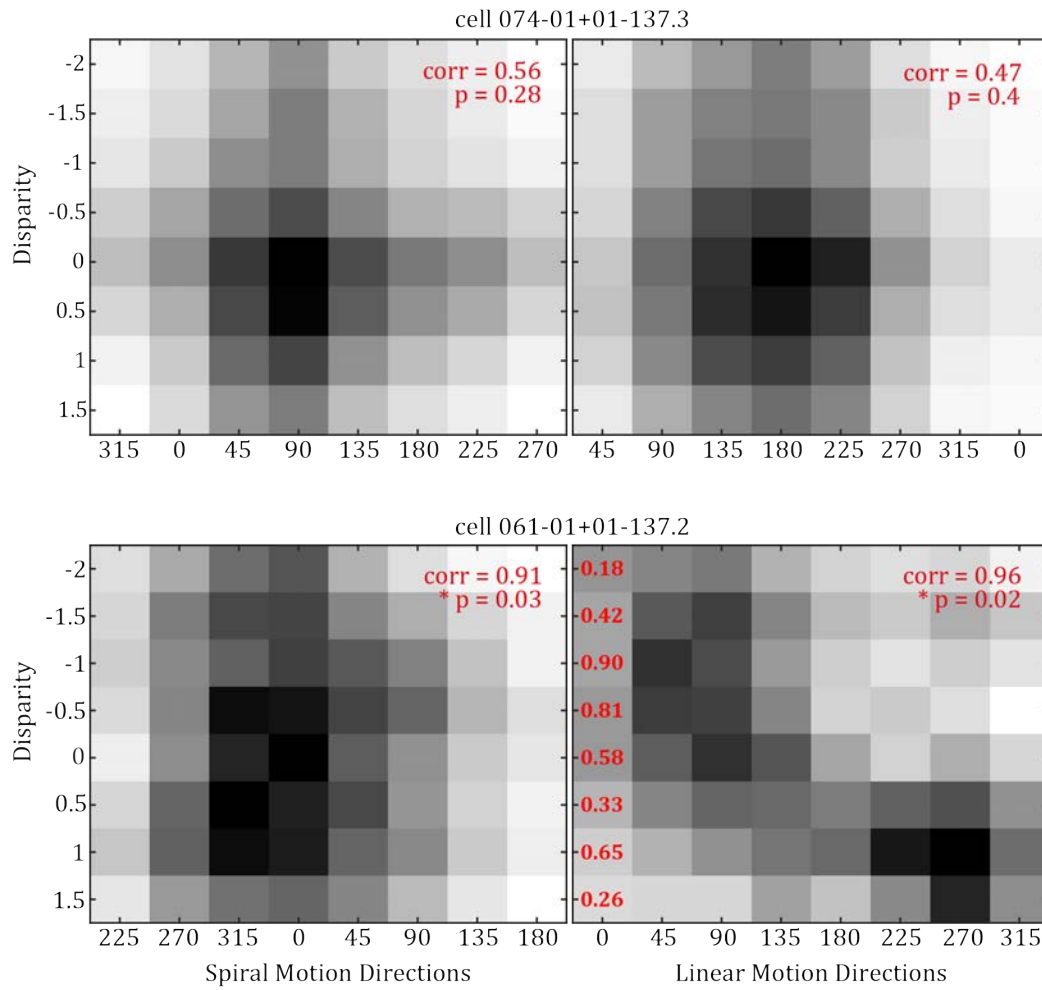


Figure 8 – Directionality and disparity selectivity joint probability heat maps for two example cells. While cell 074-01+01-137.2 (top left and right for spiral and linear motions respectively) shows no correlation between the two feature domains (Spiral $p = 0.28$, Linear $p = 0.4$), cell 061-01+01-131.2 shows a significant switch of the preferred direction together with shifts in disparity (Spiral $p = 0.03$, Linear $p = 0.02$), with linear motion accounting for the greater shift of ~ 135 degrees.

Figure 6 upper row illustrates one example unit for which there is no change of directionality together with a change in disparity, for neither spiral nor linear motion. Figure 6 lower row illustrates another example unit, displaying a shift of preferred direction for linear motion stimuli depending on the stimulus disparity. The first example summarizes the entirety of the population, making the second example the only unit showing such property. Nonetheless a closer look at the only disparity-dependent direction cell tuning, based on firing rate rather than probability, reveals a rather sudden shift of linear motion selectivity at disparity 0.5 (Figure 9). Although this behaviour is in line with previous reports (Roy et al., 1992; Takemura et al., 2001), the proportion of cells showing this response pattern in our study seems critically discordant with the aforementioned studies.

Variance explained by individual features

In order to assess the second experimental question – how much of the neuronal variability is explained by disparity, direction and their putative interaction – four generalized linear models were tested (see Methods section). To illustrate the quality of each model to describe the variability of the spike count, the resulting distributions of deviances are reported in figure 10 for the whole population of 164 cells here considered (red dots in figure 10 represent median of the population). While *disparity only* (m1) describes around 6% and 4% of the variability for spiral and linear motion respectively; *direction only* (m2) reaches 37% and 43%, making this second feature the most dominant across the population (*t-test* pairwise comparison between m2 and m1 reveals p-value < 0.001, for both spiral and linear).

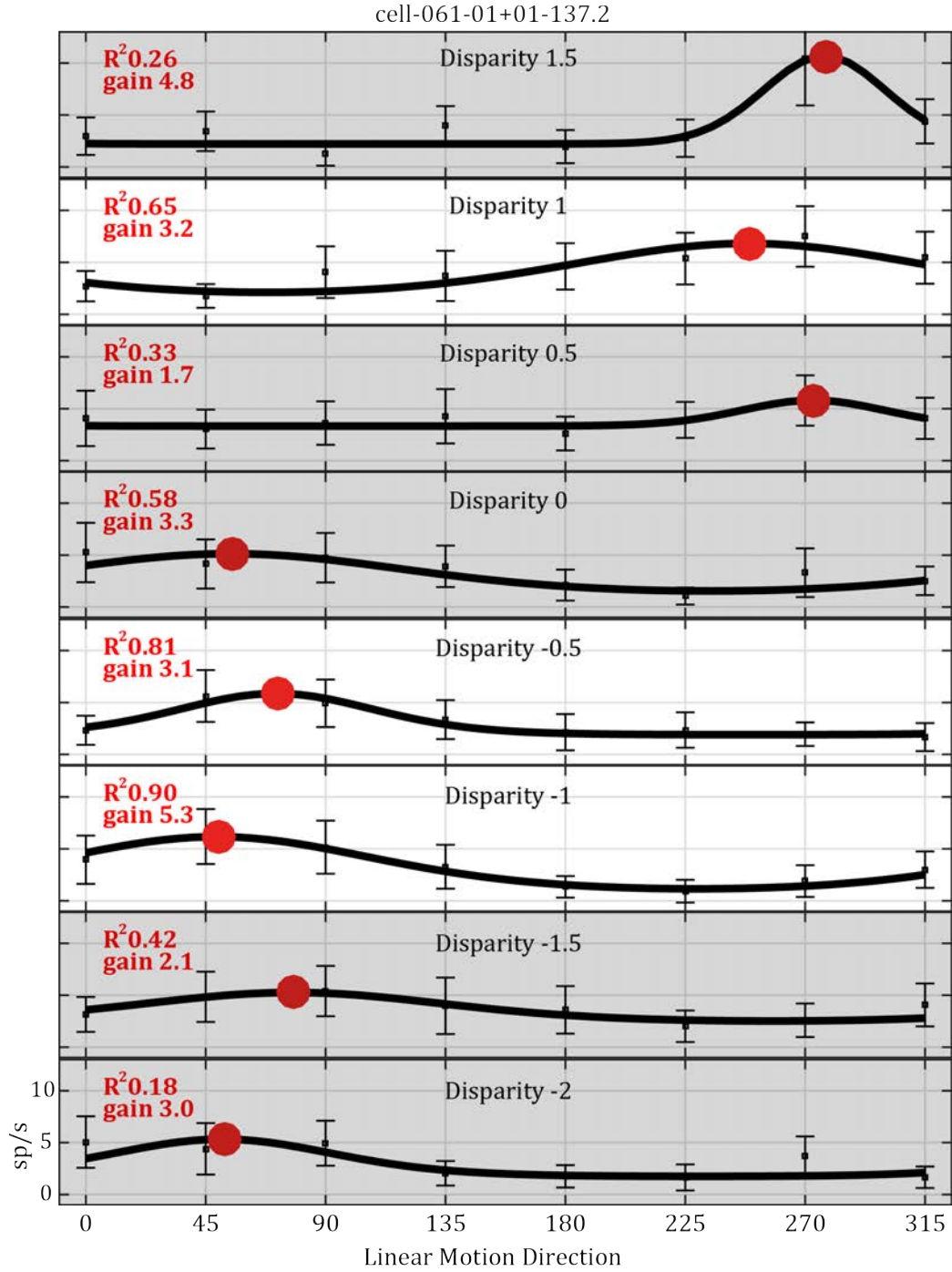


Figure 9 – Disparity dependent directional cell 061-01+01-137.2. Firing rate based tuning fit (error bars represent 95% confidence interval) at different disparity ordered from far to near and from top to bottom. Red texts indicate the resultant adjusted R^2 of the fit and the gain, as ratio between highest to lowest point of the curve. Red dots represent the preferred direction resulting from the fit. From near (negative values) to far (positive values) the preferred direction switches of ~ 135 degrees. Panels with grey background indicate those disparities at which the von Mises fit returns an adjusted r squared below our inclusion criteria.

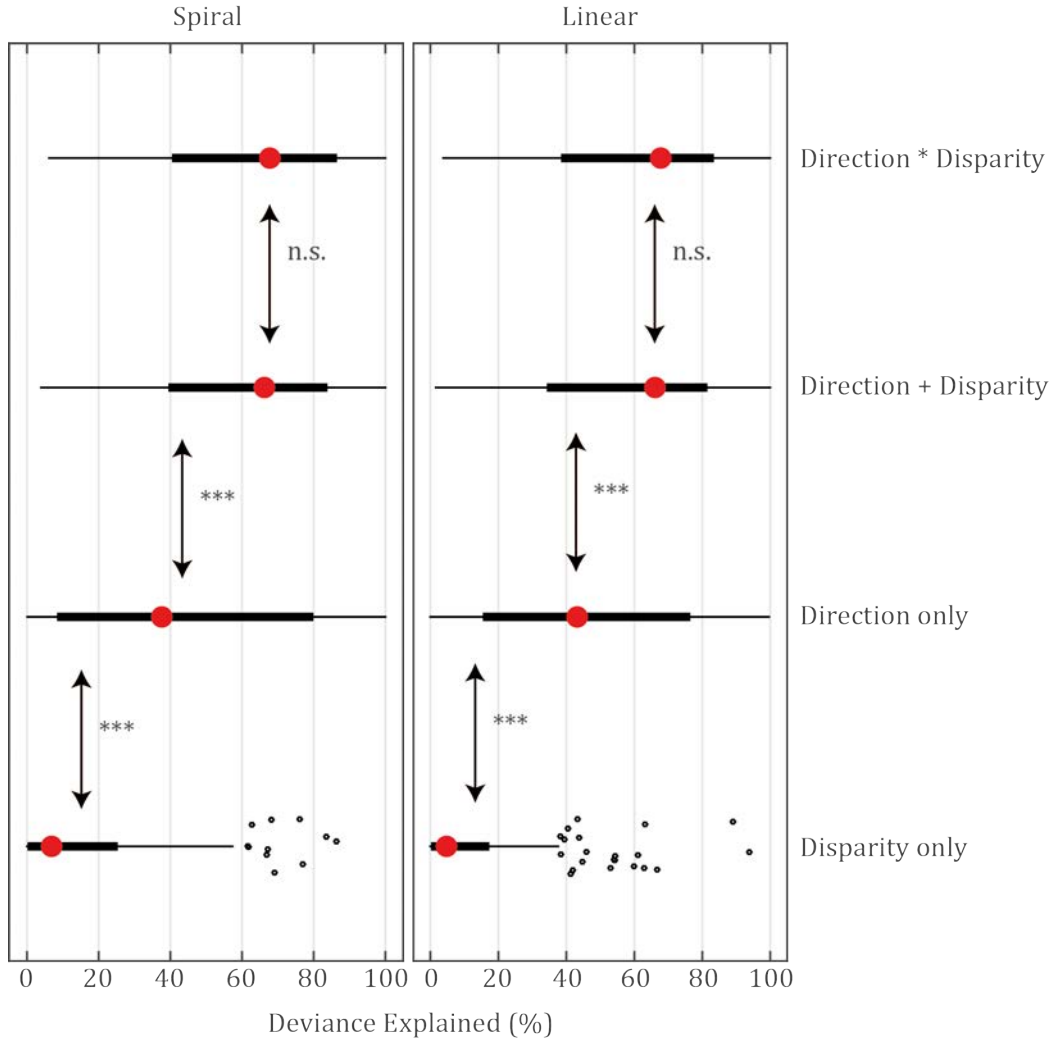


Figure 10 – Distributions of deviance explained by the four GLM models across 154 cells comprising the population in analysis. From bottom to top, each row indicates the distribution of deviances for a given model, for both spiral (left) and linear (right) motions, on the right. Red dots indicate the median of the distribution, while thick black lines indicate the 1st to 3rd quartile range, thin black lines the minimum and the maximum values of the distribution and circles the outliers. Three stars indicate p value below 0.001 to a t -test pairwise comparison and n.s. stands for non-significant difference.

Moreover, best describing the spike counts across the population are the models accounting for both feature dimensions at the same time. Model 3, considering an additive effect of direction and disparity, explains a median of 66% of the variance in both motion domains, significantly diverging from the explanatory power of both previous models (t -test pairwise comparison between m3 and m2, m3 and m1,

reveals $p\text{-value} < 0.001$, for both spiral and linear). Finally, on top of model 3, despite an interaction parameter is added in model 4, the resulting deviance of 67% for both motion domains, does not improve the explanatory power of model 3 (*t-test* pairwise comparison between m4 and m3 reveals $p\text{-value} > 0.05$, for both spiral and linear). This suggests that interaction between motion direction selectivity and disparity selectivity is not necessary to explain more variability in activity of single cells, which echoes with the PCA result, that disparity representation in the population is independent from the other features of the stimulus.

Discussion

After decades of research exploring MST's selectivity in the motion domain, two considerations seem to find ample agreement. The first one wants the middle superior temporal area to be responsible for the decomposition of optic flow information into the two major axis constituting it, namely the rotation and expansion/contraction (Duffy & Wurtz, 1991; Graziano et al., 1994; Orban et al., 1992; Saito et al., 1986; Tanaka & Saito, 1989). The second one relates to the proximity of this area to a variety of other anatomical as well as functional networks which, as a consequence, puts the area at the centre of a very diversified computational node for heading direction estimation and self-motion computation (Gu et al., 2008; Roy et al., 1992; Sakata et al., 1983; Yang et al., 2011). While this study is in substantial agreement with the first consideration, at the same time it fails to provide supporting evidences for the second. This section will first summarize the two main observations behind such dissonance and secondly will elaborate on its motives as well as consequences.

First, while almost half of the neurons here considered show tuning to binocular disparity either in the linear or in the spiral domain, in line with existing literature (Roy et al., 1992; Takemura et al., 2001; Yang et al., 2011); only 1 cell out of 164 showed disparity-dependent-direction selectivity, in striking contrast with the original 1992 study of Roy *et al* in which 40% of units there considered reversed their directionality with changes in disparity. Our proportion also fails at the comparison with a more conservative proportion of 5% DDD cells in the experiment of Yang and collaborators, 2011. The first factor one must consider when searching for potential explanations of differences in neuronal responses in higher order

visual areas, situated at the top of the *superior temporal sulcus* of the macaque brain, is the proximity with adjacent areas with similar but not identical functional properties. Additionally, in the case of MST one must also take into consideration that such area has been partitioned in several sub regions with rather diversified properties (Desimone & Ungerleider, 1986). For example, while showing very clear sensitivity to binocular disparity, no DDD cells were found in area MT (Smolyanskaya et al., 2013) and yet a very different sensitivity to disparity was described for area MSTl (Eifuku & Wurtz, 1999) compared to MSTd. Moreover, when comparing studies in which area localization was done histologically, with studies whose localization was MRT based (like in our case), different cellular responses can simply be due to having probed unidentified sub portions of the target area. Finally, while the discretisation of brain regions (Brodmann, 1909) sits at the very core of most neurophysiology works, one must consider that in reality, brain areas gradually fade into one another and that borders strongly depend on the methodology and the statistics employed (Coalson et al., 2016).

The second major observation resulting from this study relates to the explanatory power of the two visual features here considered, motion direction and binocular disparity. Under the hypothesis that MST is indeed the node in which joint selectivity of depth and motion is computed and passed to next hierarchical processing stages, one would expect to find that complex interaction of the aforementioned features would significantly better explain the spiking behaviour of the population. On the contrary, considering that the *interaction* model did not better explain the distribution of spikes count than the *additive* model, our generalized linear model approach suggests independence of the two features at the

MST spiking level. Making disparity and direction independently accessible to later computational stages, brings MST's functionality closer to MT (DeAngelis & Newsome, 1999) and at the same time moves some of the numerous functional responsibilities of this area to later computational nodes (Raffi, Persiani, Piras, & Squatrito, 2014).

A number of studies has found disparity sensitive neurons in MT (DeAngelis & Newsome, 1999; Maunsell & Van Essen, 1983b; Smolyanskaya et al., 2013) and MST (Roy et al., 1992; Smolyanskaya et al., 2013; but see Yang et al., 2011). However, with our principal component analysis we found that the population decoder for disparity is more related to the coarse categorization of near space and far space rather than a continuous representation of depths. While some studies have found that MST neurons selectively respond to stimuli at different disparities, there has been no decisive conclusion on the role of the area MST in the perception of depth. As our analysis suggest, the contribution from MST to an accurate and fine perception of depth is probably very limited.

In conclusion, we confirmed that motion direction and binocular disparity are both represented in MST. However, individual cells and linear population decoding analyses both showed that the encoding of binocular disparity is independent from the other features of the stimuli. Therefore, while it seems reasonable that cells with disparity dependent direction tuning could be crucial for self-motion perception, such a integration may not happen directly in MST. Further investigations in areas higher in the visual hierarchy are necessary to finally reveal the underlying mechanism for self-motion perception.

References

- Bair, W., Cavanaugh, J. R., Smith, M. A., & Movshon, J. A. (2002). The timing of response onset and offset in macaque visual neurons. *Journal of Neuroscience*, 22(8), 3189–3205.
- Berens, P. (2008). Comparing the feature selectivity of the gamma-band of the local field potential and the underlying spiking activity in primate visual cortex. *Frontiers in Systems Neuroscience*, 2, 1–11. <http://doi.org/10.3389/neuro.06.002.2008>
- Borghuis, B. G., Perge, J. A., Vajda, I., van Wezel, R. J. A., van de Grind, W. A., & Lankheet, M. J. M. (2003). The motion reverse correlation (MRC) method. *Journal of Neuroscience Methods*, 123(2), 153–166. [http://doi.org/10.1016/S0165-0270\(02\)00347-3](http://doi.org/10.1016/S0165-0270(02)00347-3)
- Britten, K. H., & van Wezel, R. J. A. (2002). Area MST and Heading Perception in Macaque Monkeys. *Cerebral Cortex*, 12(7), 692–701. <http://doi.org/10.1093/cercor/12.7.692>
- Brodmann, K. (1909). Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues.
- Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network (Bristol, England)*, 12(2), 199–213.
- Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 1–11. <http://doi.org/10.1038/nature18933>
- de Boer, R., & Kuypers, P. (1968). Triggered correlation. *IEEE Transactions on Bio-Medical Engineering*, 15(3), 169–179.
- DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *The Journal of Neuroscience*, 19(4), 1398–1415.
- Desimone, R., & Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *Journal of Comparative Neurology*, 248(2), 164–189. <http://doi.org/10.1002/cne.902480203>
- Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6), 1329–1345. <http://doi.org/10.1152/jn.00847.2015>
- Eifuku, S., & Wurtz, R. H. (1999). Response to motion in extrastriate area MSTl: disparity sensitivity. *Journal of Neurophysiology*, 82(5), 2462–2475. <http://doi.org/10.1146/annurev.ne.08.030185.002203>
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1), 1–47. <http://doi.org/10.1093/cercor/1.1.1>
- Graziano, M. S., Andersen, R. A., & Snowden, R. J. (1994). Tuning of MST neurons to spiral motions. *The Journal of Neuroscience*, 14(1), 54–67.
- Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11(10), 1201–1210. <http://doi.org/10.1038/nn.2191>
- Kawano, K., Inoue, Y., Takemura, A., Kodaka, Y., & Miles, F. A. (1999). The role of MST neurons during ocular tracking in 3D space. *International Review of Neurobiology*, 44, 49–63.
- Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network (Bristol, England)*, 9(4), R53–

- Maunsell, J. H., & Van Essen, D. C. (1983a). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49(5), 1127–1147.
- Maunsell, J. H., & Van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, 49(5), 1148–1167.
- Maunsell, J. H., & Van Essen, D. C. (1983c). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neuroscience*, 3(12), 2563–2586.
- Mendoza-Halliday, D., Torres, S., & Martinez-Trujillo, J. C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nature Publishing Group*, 17(9), 1255–1262. <http://doi.org/10.1038/nm.3785>
- Mineault, P. J., Khawaja, F. A., & Butts, D. A. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 109, pp. E972–E980). Proceedings of the <http://doi.org/10.1073/pnas.1115685109/-/DCSupplemental>
- Orban, G. A., Lagae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H., & Torre, V. (1992). First-order analysis of optical flow in monkey brain. *Proceedings of the National Academy of Sciences*, 89(7), 2595–2599.
- Perry, C. J., & Fallah, M. (2014). Feature integration and object representations along the dorsal stream visual hierarchy. *Front Comput Neurosci*. <http://doi.org/10.3389/fncom.2014.00084/abstract>
- Price, N. S. C., & Born, R. T. (2013). Adaptation to Speed in Macaque Middle Temporal and Medial Superior Temporal Areas. *The Journal of Neuroscience*, 33(10), 4359–4368. <http://doi.org/10.1523/JNEUROSCI.3165-12.2013>
- Raffi, M., Persiani, M., Piras, A., & Squatrito, S. (2014). Optic flow neurons in area PEc integrate eye and head position signals. *Neuroscience Letters*, 568, 23–28. <http://doi.org/10.1016/j.neulet.2014.03.042>
- Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, 387(6630), 281–284. <http://doi.org/10.1038/387281a0>
- Roy, J. P., & Wurtz, R. H. (1990). The role of disparity-sensitive cortical neurons in signalling the direction of self-motion. *Nature*, 348(6297), 160–162. <http://doi.org/10.1038/348160a0>
- Roy, J. P., Komatsu, H., & Wurtz, R. H. (1992). Disparity sensitivity of neurons in monkey extrastriate area MST. *The Journal of Neuroscience*, 12(7), 2478–2492.
- Saito, H.-A., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., & Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *The Journal of Neuroscience*, 6(1), 145–157.
- Sakata, H., Shibutani, H., & Kawano, K. (1983). Functional properties of visual tracking neurons in posterior parietal association cortex of the monkey. *Journal of Neurophysiology*, 49(6), 1364–1380. <http://doi.org/10.1152/jn.00847.2015>
- Smolyanskaya, A., Ruff, D. A., & Born, R. T. (2013). Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable. *Journal of Neurophysiology*, 110(12), 2806–2816. <http://doi.org/10.1152/jn.00573.2013>
- Takahashi, K., Gu, Y., May, P. J., Newlands, S. D., DeAngelis, G. C., & Angelaki, D. E.

- (2007). Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. *Journal of Neuroscience*, 27(36), 9742–9756. <http://doi.org/10.1523/JNEUROSCI.0817-07.2007>
- Takemura, A., Inoue, Y., Kawano, K., Quaia, C., & Miles, F. A. (2001). Single-unit activity in cortical area MST associated with disparity-vergence eye movements: evidence for population coding. *Journal of Neurophysiology*, 85(5), 2245–2266.
- Tanaka, K., & Saito, H.-A. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, 62(3), 626–641.
- Tanaka, K., Sugita, Y., Moriya, M., & Saito, H. (1993). Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *Journal of Neurophysiology*, 69(1), 128–142.
- Tchernikov, I., & Fallah, M. (2010). A Color Hierarchy for Automatic Target Selection. *PLoS ONE*, 5(2), e9338–4. <http://doi.org/10.1371/journal.pone.0009338>
- Treue, S., & Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, 382(6591), 539–541. <http://doi.org/10.1038/382539a0>
- Yang, Y., Liu, S., Chowdhury, S. A., DeAngelis, G. C., & Angelaki, D. E. (2011). Binocular disparity tuning and visual-vestibular congruency of multisensory neurons in macaque parietal cortex. *Journal of Neuroscience*, 31(49), 17905–17916. <http://doi.org/10.1523/JNEUROSCI.4032-11.2011>
- Zeki, S. M. (1978a). Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, Vol. 274, pp 423-428
- Zeki, S. M. (1978b). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *The Journal of Physiology*, 277, 273–290. [http://doi.org/10.1111/\(ISSN\)1469-7793](http://doi.org/10.1111/(ISSN)1469-7793)

Supplementary material

List of cells included in the analysis, collected through a recording chamber placed on the left hemisphere of monkey Igg (-3.1 mm AP and -19.5 mm ML), and a recording chamber placed on the right hemisphere of monkey Nic (-2.6mm AP and 14.0mm ML)

Cell ID	date	x	y	z (micron)
'igg-002-01+01-129.1'	150407	1	-4	4323
'igg-002-01+01-129.2'	150407	1	-4	4323
'igg-004-01+01-129.1'	150410	-3	-2	12391
'igg-004-01+01-129.2'	150410	-3	-2	12391
'igg-006-01+01-129.1'	150415	-3	-5	10750
'igg-011-01+01-129.1'	150423	-4	0	7086
'igg-011-01+01-129.2'	150423	-4	0	7086
'igg-013-01+01-129.1'	150428	-5	1	10727
'igg-013-01+01-129.3'	150428	-5	1	10727
'igg-014-01+01-129.1'	150429	-3	-1	6149
'igg-014-01+01-129.2'	150429	-3	-1	6149
'igg-015-01+01-129.1'	150504	-4	1	10293
'igg-015-01+01-129.2'	150504	-4	1	10293
'igg-016-01+01-129.1'	150505	-4	1	9374
'igg-018-01+01-129.1'	150513	-4	1	11169
'igg-019-01+01-129.1'	150519	-4	1	8500
'igg-019-01+01-129.2'	150519	-4	1	8500
'igg-020-01+01-129.1'	150520	-4	1	8400
'igg-021-01+01-129.1'	150525	-4	1	8750
'igg-023-01+01-129.1'	150528	-4	1	7930
'igg-024-01+01-129.1'	150529	-4	1	9187
'igg-025-01+01-129.1'	150608	-4	1	6844
'igg-025-01+01-129.2'	150608	-4	1	6844
'igg-026-01+01-129.1'	150609	-4	1	9834
'igg-027-01+01-130.1'	150618	-4	1	11782
'igg-028-01+01-129.1'	150622	-4	1	12666
'igg-028-01+01-129.2'	150622	-4	1	12666
'igg-028-01+01-130.2'	150622	-4	1	12666
'igg-029-01+01-129.1'	150623	-4	1	12500
'igg-029-01+01-129.2'	150623	-4	1	12500
'igg-029-01+01-129.3'	150623	-4	1	12500
'igg-032-01+01-129.1'	150728	-5	0	8311
'igg-032-01+01-129.2'	150728	-5	0	8311
'igg-033-01+01-129.1'	150729	-5	0	8256
'igg-033-01+01-129.2'	150729	-5	0	8256
'igg-033-01+01-129.3'	150729	-5	0	8256
'igg-034-01+01-129.1'	150730	-5	0	8250
'igg-034-01+01-129.4'	150730	-5	0	8250
'igg-035-01+01-129.1'	150731	-5	0	8550
'igg-035-01+01-129.2'	150731	-5	0	8550
'igg-035-01+01-129.3'	150731	-5	0	8550
'igg-036-01+01-129.1'	150803	-5	0	8550
'igg-037-01+01-129.1'	150804	-5	-1	7530
'igg-037-01+01-129.2'	150804	-5	-1	7530
'igg-037-01+01-129.3'	150804	-5	-1	7530
'igg-037-01+01-129.4'	150804	-5	-1	7530
'igg-037-01+01-130.1'	150804	-5	-1	7680
'igg-037-01+01-130.5'	150804	-5	-1	7680
'igg-039-01+01-129.1'	150806	-5	-1	8288,00
'igg-039-01+01-130.1'	150806	-5	-1	8005,00
'igg-040-01+01-129.1'	150810	-5	-1	9370,00
'igg-041-01+01-130.1'	150812	-5	-1	7222,00
'igg-042-01+01-129.1'	150813	-5	-1	7258,00

'igg-042-01+01-130.1'	150813	-5	-1	7346,00
'igg-042-01+01-130.2'	150813	-5	-1	7346,00
'igg-045-01+01-129.1'	151020	-5	-1	7500,00
'igg-045-01+01-129.2'	151020	-5	-1	7500,00
'igg-046-01+01-129.1'	151022	-5	-1	8724,00
'igg-046-01+01-129.2'	151022	-5	-1	8724,00
'igg-047-01+01-129.1'	151023	-5	-1	5067,00
'igg-047-01+01-129.2'	151023	-5	-1	5067,00
'igg-047-01+01-129.3'	151023	-5	-1	5067,00
'igg-048-01+01-129.1'	151027	-5	-1	5330,00
'igg-049-01+01-129.1'	151110	-5	-1	9612,00
'igg-049-01+01-129.2'	151110	-5	-1	9612,00
'igg-050-01+01-129.1'	151111	-5	-1	8635,00
'igg-050-01+01-129.2'	151111	-5	-1	8635,00
'igg-051-01+01-129.1'	151112	-5	-1	6742,00
'igg-053-01+01-129.1'	151119	-5	-1	4510,00
'igg-053-01+01-129.2'	151119	-5	-1	4510,00
'igg-053-01+01-133.1'	151119	-5	-1	5360,00
'igg-053-01+01-133.2'	151119	-5	-1	5360,00
'igg-054-01+01-133.1'	151120	-5	-1	5380,00
'igg-054-01+01-133.2'	151120	-5	-1	5380,00
'igg-054-02+01-129.1'	151120	-5	-1	4909,00
'igg-054-02+01-129.2'	151120	-5	-1	4909,00
'igg-054-02+01-129.3'	151120	-5	-1	4909,00
'igg-054-02+01-133.1'	151120	-5	-1	5691,00
'igg-054-02+01-133.2'	151120	-5	-1	5691,00
'igg-054-02+01-133.3'	151120	-5	-1	5691,00
'igg-055-01+01-129.1'	151123	-5	-1	5958,00
'igg-055-01+01-129.2'	151123	-5	-1	5958,00
'igg-055-01+01-133.1'	151123	-5	-1	6234,00
'igg-055-01+01-133.2'	151123	-5	-1	6234,00
'igg-055-01+01-133.3'	151123	-5	-1	6234,00
'igg-056-01+01-129.1'	151124	-5	-1	6500,00
'igg-056-01+01-129.2'	151124	-5	-1	6500,00
'igg-056-01+01-133.1'	151124	-5	-1	6060,00
'igg-056-01+01-133.2'	151124	-5	-1	6060,00
'igg-056-01+01-133.3'	151124	-5	-1	6060,00
'igg-057-01+01-129.1'	151125	-5	-1	4321,00
'igg-057-01+01-129.2'	151125	-5	-1	4321,00
'igg-057-01+01-129.3'	151125	-5	-1	4321,00
'igg-057-01+01-133.1'	151125	-5	-1	1825,00
'igg-058-01+01-129.1'	151130	-5	-1	3839,00
'igg-058-01+01-129.2'	151130	-5	-1	3839,00
'igg-058-01+01-129.3'	151130	-5	-1	3839,00
'igg-058-01+01-129.4'	151130	-5	-1	3839,00
'igg-058-01+01-129.5'	151130	-5	-1	3839,00
'igg-058-01+01-133.1'	151130	-5	-1	3145,00
'igg-058-01+01-133.2'	151130	-5	-1	3145,00
'igg-058-01+01-133.3'	151130	-5	-1	3145,00
'igg-059-01+01-129.1'	151207	-5	-1	3839,00
'igg-059-01+01-129.2'	151207	-5	-1	3839,00
'igg-059-01+01-133.1'	151207	-5	-1	3145,00
'igg-059-01+01-133.2'	151207	-5	-1	3145,00
'igg-059-01+01-133.3'	151207	-5	-1	3145,00

'igg-060-01+01-129.1'	151208	-5	-1	5052,00
'igg-060-01+01-129.2'	151208	-5	-1	5052,00
'igg-060-01+01-129.3'	151208	-5	-1	5052,00
'igg-060-01+01-133.1'	151208	-5	-1	9928,00
'igg-060-01+01-137.1'	151208	-5	-1	4725,00
'igg-060-01+01-137.2'	151208	-5	-1	4725,00
'igg-060-01+01-137.3'	151208	-5	-1	4725,00
'igg-061-01+01-133.1'	151210	-5	-1	9441,00
'igg-061-01+01-133.2'	151210	-5	-1	9441,00
'igg-061-01+01-137.1'	151210	-5	-1	5161,00
'igg-061-01+01-137.2'	151210	-5	-1	5161,00
'igg-062-01+01-129.1'	151216	-5	-2	2500,00
'igg-062-01+01-129.2'	151216	-5	-2	2500,00
'igg-062-01+01-129.3'	151216	-5	-2	2500,00
'igg-062-01+01-129.4'	151216	-5	-2	2500,00
'igg-062-01+01-133.1'	151216	-5	-2	8000,00
'igg-062-01+01-133.2'	151216	-5	-2	8000,00
'igg-062-01+01-137.1'	151216	-5	-2	4500,00
'igg-062-01+01-137.2'	151216	-5	-2	4500,00
'igg-062-01+01-137.3'	151216	-5	-2	4500,00
'igg-062-01+01-137.4'	151216	-5	-2	4500,00
'igg-062-01+01-137.5'	151216	-5	-2	4500,00
'igg-062-01+01-137.6'	151216	-5	-2	4500,00
'igg-062-02+01-129.1'	151216	-5	-2	8058,00
'igg-062-02+01-129.2'	151216	-5	-2	8058,00
'igg-062-02+01-129.3'	151216	-5	-2	8058,00
'igg-062-02+01-129.4'	151216	-5	-2	8058,00
'igg-062-02+01-133.1'	151216	-5	-2	8200,00
'igg-062-02+01-133.2'	151216	-5	-2	8200,00
'igg-063-01+01-129.1'	151217	-5	-2	3500,00
'igg-063-01+01-129.2'	151217	-5	-2	3500,00
'igg-063-01+01-129.3'	151217	-5	-2	3500,00
'igg-063-01+01-129.7'	151217	-5	-2	3500,00
'igg-063-01+01-129.9'	151217	-5	-2	3500,00
'igg-063-01+01-133.1'	151217	-5	-2	577,00
'igg-063-01+01-133.2'	151217	-5	-2	577,00
'igg-063-01+01-137.1'	151217	-5	-2	4588,00
'igg-065-01+01-129.1'	151230	-5	-2	8966,00
'igg-065-01+01-129.2'	151230	-5	-2	8966,00
'igg-065-01+01-129.3'	151230	-5	-2	8966,00
'igg-065-01+01-129.4'	151230	-5	-2	8966,00
'igg-065-01+01-129.5'	151230	-5	-2	8966,00
'igg-065-01+01-129.6'	151230	-5	-2	8966,00
'igg-065-01+01-133.1'	151230	-5	-2	9509,00
'igg-065-01+01-133.2'	151230	-5	-2	9509,00
'igg-066-01+01-129.1'	160104	-5	-2	3742,00
'igg-066-01+01-129.2'	160104	-5	-2	3742,00
'igg-066-01+01-129.3'	160104	-5	-2	3742,00
'igg-066-01+01-129.6'	160104	-5	-2	3742,00
'igg-066-01+01-133.1'	160104	-5	-2	6889,00
'igg-066-01+01-133.2'	160104	-5	-2	6889,00
'igg-066-01+01-133.3'	160104	-5	-2	6889,00
'igg-067-01+01-129.1'	160105	-5	-1	8512,00
'igg-067-01+01-129.2'	160105	-5	-1	8512,00
'igg-067-01+01-129.3'	160105	-5	-1	8512,00
'igg-067-01+01-129.4'	160105	-5	-1	8512,00
'igg-067-01+01-129.5'	160105	-5	-1	8512,00
'igg-067-01+01-133.1'	160105	-5	-1	4522,00
'igg-068-01+01-129.1'	160106	-5	-1	9005,00
'igg-068-01+01-129.2'	160106	-5	-1	9005,00
'igg-068-01+01-129.4'	160106	-5	-1	9005,00
'igg-068-01+01-129.5'	160106	-5	-1	9005,00

'igg-068-01+01-129.6'	160106	-5	-1	9005,00
'igg-068-01+01-129.7'	160106	-5	-1	9005,00
'igg-068-01+01-129.8'	160106	-5	-1	9005,00
'igg-068-01+01-133.1'	160106	-5	-1	4300,00
'igg-068-01+01-133.5'	160106	-5	-1	4300,00
'igg-068-01+01-133.6'	160106	-5	-1	4300,00
'igg-069-01+01-129.1'	160107	-5	-1	2745,00
'igg-070-01+01-129.1'	160111	-5	-1	2877,00
'igg-070-01+01-133.1'	160111	-5	-1	2673,00
'igg-070-01+01-137.1'	160111	-5	-1	3971,00
'igg-070-02+01-133.1'	160111	-5	-1	3560,00
'igg-070-02+01-133.2'	160111	-5	-1	3560,00
'igg-070-02+01-137.1'	160111	-5	-1	4760,00
'igg-071-01+01-133.2'	160113	-5	-1	3010,00
'igg-071-01+01-137.1'	160113	-5	-1	3420,00
'igg-071-01+01-137.2'	160113	-5	-1	3420,00
'igg-071-02+01-133.1'	160113	-5	-1	3570,00
'igg-071-02+01-133.2'	160113	-5	-1	3570,00
'igg-071-02+01-137.1'	160113	-5	-1	4145,00
'igg-071-02+01-137.2'	160113	-5	-1	4145,00
'igg-071-02+01-137.3'	160113	-5	-1	4145,00
'igg-072-01+01-129.1'	160114	-5	-1	9520,00
'igg-072-01+01-137.1'	160114	-5	-1	4852,00
'igg-072-01+01-137.2'	160114	-5	-1	4852,00
'igg-072-01+01-137.4'	160114	-5	-1	4852,00
'igg-072-01+01-137.6'	160114	-5	-1	4852,00
'igg-073-01+01-129.1'	160119	-5	-1	9122,00
'igg-073-01+01-137.1'	160119	-5	-1	9100,00
'igg-073-01+01-137.2'	160119	-5	-1	9100,00
'igg-074-01+01-133.1'	160121	-5	-1	8465,00
'igg-074-01+01-133.3'	160121	-5	-1	8465,00
'igg-074-01+01-137.1'	160121	-5	-1	9000,00
'igg-074-01+01-137.2'	160121	-5	-1	9000,00
'igg-074-01+01-137.3'	160121	-5	-1	9000,00
'igg-074-01+01-137.4'	160121	-5	-1	9000,00
'igg-075-01+01-129.1'	160217	-5	0	2500,00
'igg-075-01+01-129.4'	160217	-5	0	2500,00
'igg-075-01+01-129.5'	160217	-5	0	2500,00
'igg-075-01+01-133.1'	160217	-5	0	4234,00
'igg-075-01+01-133.2'	160217	-5	0	4234,00
'igg-075-01+01-133.3'	160217	-5	0	4234,00
'igg-075-01+01-133.4'	160217	-5	0	4234,00
'igg-075-01+01-133.5'	160217	-5	0	4234,00
'igg-075-01+01-133.6'	160217	-5	0	4234,00
'igg-075-01+01-133.7'	160217	-5	0	4234,00
'igg-076-01+01-133.1'	160219	-5	0	5128,00
'igg-076-01+01-137.1'	160219	-5	0	5679,00
'igg-076-01+01-137.2'	160219	-5	0	5679,00
'igg-076-01+01-137.3'	160219	-5	0	5679,00
'igg-077-01+01-133.2'	160223	-5	0	2100,00
'igg-077-01+01-137.3'	160223	-5	0	5230,00
'nic-016-04+01-17.1'	160809	4	2	12879,00
'nic-016-04+01-17.2'	160809	4	2	14014,00
'nic-017-01+01-17.1'	160810	3	2	13731,00
'nic-017-01+01-17.2'	160810	3	2	13731,00
'nic-017-02+01-17.1'	160810	3	2	14014,00
'nic-017-02+01-17.2'	160810	3	2	14014,00
'nic-018-03+01-17.1'	160811	3	2	12814,00
'nic-018-03+01-17.2'	160811	3	2	12814,00
'nic-019-02+01-17.1'	160812	5	2	11137,00
'nic-019-02+01-17.2'	160812	5	2	11137,00
'nic-019-03+01-17.1'	160812	4	2	18000,00

Excluded cells due to insufficient number of repetitions or no response to visual stimulation:

'igg-001-01+01-129.1'
'igg-002-01+01-129.1'
'igg-002-01+01-129.2'
'igg-004-01+01-129.1'
'igg-004-01+01-129.2'
'igg-006-01+01-129.1'
'igg-011-01+01-129.1'
'igg-011-01+01-129.2'
'igg-013-01+01-129.1'
'igg-013-01+01-129.3'
'igg-014-01+01-129.1'
'igg-015-01+01-129.1'
'igg-015-01+01-129.2'
'igg-016-01+01-129.1'
'igg-018-01+01-129.1'
'igg-019-01+01-129.1'
'igg-019-01+01-129.2'
'igg-020-01+01-129.1'
'igg-023-01+01-129.1'
'igg-024-01+01-129.1'
'igg-025-01+01-129.2'
'igg-028-01+01-129.2'
'igg-029-01+01-129.2'
'igg-033-01+01-129.1'

'igg-033-01+01-129.2'
'igg-033-01+01-129.3'
'igg-034-01+01-129.2'
'igg-034-01+01-129.4'
'igg-035-01+01-129.1'
'igg-035-01+01-129.2'
'igg-037-01+01-129.1'
'igg-037-01+01-129.2'
'igg-037-01+01-129.3'
'igg-037-01+01-130.3'
'igg-039-01+01-129.1'
'igg-040-01+01-129.1'
'igg-041-01+01-130.1'
'igg-046-01+01-129.2'
'igg-047-01+01-129.3'
'igg-049-01+01-129.1'
'igg-049-01+01-129.2'
'igg-057-01+01-129.1'
'igg-058-01+01-129.5'
'igg-058-01+01-133.3'
'igg-059-01+01-129.1'
'igg-061-01+01-133.1'
'igg-061-01+01-133.2'

'igg-061-01+01-137.3'
'igg-062-01+01-129.1'
'igg-062-01+01-129.2'
'igg-062-01+01-129.3'
'igg-062-01+01-129.4'
'igg-062-01+01-137.2'
'igg-062-01+01-137.3'
'igg-062-01+01-137.4'
'igg-062-02+01-129.4'
'igg-066-01+01-129.4'
'igg-066-01+01-129.5'
'igg-066-01+01-129.7'
'igg-066-01+01-133.4'
'igg-067-01+01-133.2'
'igg-067-01+01-133.3'
'igg-067-01+01-133.4'
'igg-069-01+01-129.2'
'igg-070-01+01-129.2'
'igg-070-01+01-129.3'
'igg-070-01+01-129.4'
'igg-070-01+01-129.5'
'igg-070-01+01-133.2'
'igg-070-01+01-133.3'
'igg-070-02+01-129.1'

'igg-070-02+01-129.2'
'igg-073-01+01-129.2'
'igg-074-01+01-133.2'
'igg-075-01+01-129.2'
'igg-075-01+01-129.3'
'igg-075-01+01-129.6'
'igg-076-01+01-129.1'
'igg-076-01+01-129.3'
'igg-076-01+01-129.4'
'igg-076-01+01-133.2'
'igg-077-01+01-129.1'
'igg-077-01+01-129.2'
'igg-077-01+01-137.1'
'igg-078-01+01-129.1'
'igg-078-01+01-129.2'
'igg-078-01+01-129.3'
'igg-078-01+01-129.4'
'igg-078-01+01-133.1'
'igg-078-01+01-133.2'
'igg-078-01+01-133.3'
'igg-078-01+01-133.4'
'igg-078-01+01-133.5'
'igg-078-01+01-137.2'
'igg-078-01+01-137.3'



ORIGINAL ARTICLE

Spatial Attention Reduces Burstiness in Macaque Visual Cortical Area MST

Cheng Xue¹, Daniel Kaping^{1,2}, Sonia Baloni Ray^{1,3}, B. Suresh Krishna^{1,†}, and Stefan Treue^{1,4,5,†}

¹Cognitive Neuroscience Laboratory, German Primate Center, Goettingen 37077, Germany, ²Experimental Neurobiology, National Institute of Mental Health, Klecany 25067, Czech Republic, ³Centre of Behavioural and Cognitive Sciences, University of Allahabad, Allahabad 211001, UP, India, ⁴Faculty of Biology and Psychology, Goettingen University, Goettingen 37073, Germany and ⁵Leibniz Science Campus Primate Cognition, Goettingen 37073, Germany

Address correspondence to Cheng Xue. Email: cxue@gwdg.de; B. Suresh Krishna. Email: skrishna@dpz.eu; Stefan Treue. Email: treue@gwdg.de.

[†]Joint senior authors.

Abstract

Visual attention modulates the firing rate of neurons in many primate cortical areas. In V4, a cortical area in the ventral visual pathway, spatial attention has also been shown to reduce the tendency of neurons to fire closely separated spikes (burstiness). A recent model proposes that a single mechanism accounts for both the firing rate enhancement and the burstiness reduction in V4, but this has not been empirically tested. It is also unclear if the burstiness reduction by spatial attention is found in other visual areas and for other attentional types. We therefore recorded from single neurons in the medial superior temporal area (MST), a key motion-processing area along the dorsal visual pathway, of two rhesus monkeys while they performed a task engaging both spatial and feature-based attention. We show that in MST, spatial attention is associated with a clear reduction in burstiness that is independent of the concurrent enhancement of firing rate. In contrast, feature-based attention enhances firing rate but is not associated with a significant reduction in burstiness. These results establish burstiness reduction as a widespread effect of spatial attention. They also suggest that in contrast to the recently proposed model, the effects of spatial attention on burstiness and firing rate emerge from different mechanisms.

Key words: attention, burstiness, monkey neurophysiology, visual cortex

Introduction

Attention is a critical component of sensory processing in organisms ranging from insects to humans (Carrasco 2011; Wiederman and O'Carroll 2013). It serves to preferentially allocate sparse processing resources to currently relevant sensory input, thereby privileging it over the remaining inputs. In humans and other primates, visual attention enhances the processing of task-relevant spatial locations and visual features (such as a particular motion direction or color) that leads to improved visual performance at

these spatial locations and features (Desimone and Duncan 1995; Moore and Armstrong 2003; Bichot et al. 2015). The perceptual improvements induced by spatial and feature-based attention are accompanied by a range of neural effects that affect neuronal spike-rate (Desimone and Duncan 1995; Treue 2001; Bisley 2011), the temporal patterning of spike trains (Anderson et al. 2013), the mutual correlation between neurons (Cohen and Maunsell 2009, 2011a, 2011b; Mitchell et al. 2009) and the local field potential (Fries 2009; Esghaei et al. 2015). These effects have been hypothesized to

improve the sensory representation of attended stimuli by enhancing neural responses and reducing noise among neurons that represent the attended locations and/or features. Recently, it has been shown in V4, a key locus in the ventral stream of visual cortical information processing, that attention can also modulate aspects of neuronal firing patterns that operate on a fast time-scale: burstiness, defined as the tendency of a neuron to discharge consecutive spikes at very short inter-spike intervals, decreases in the broad-spiking neurons of area V4 when spatial attention is directed into their receptive field (RF) (Anderson et al. 2013). Though the specific functional consequence of this attentional modulation remains unknown, the effect is intriguing, because the functional properties and neural utility of bursts in spike trains have been a topic of much speculation and interest (Bair et al. 1994; Krahe and Gabbiani 2004; Izhikevich 2007). A current and plausible hypothesis states that bursts enhance information transfer because neuronal inputs composed of closely spaced spikes are more efficient at driving postsynaptic neurons which act as coincidence detectors because of their short integration time constants (Lisman 1997). As pointed out by Anderson et al. (2013), this hypothesis predicts that to drive downstream neurons more efficiently, burstiness would increase when attention is directed towards a neuron's RF. However, the burstiness reduction observed indicates the opposite.

At present, it remains unclear if the effect of spatial attention on burstiness is restricted to the ventral pathway or even only V4 and whether it extends to other types of attention. Furthermore, though it has been recently proposed based on a computational model that the effects of spatial attention on burstiness and firing rate emerge from a common mechanism (Anderson et al. 2013), there is no empirical data on how the attentional modulation of burstiness relates to the well-known modulation of firing rate by attention. To address this, we performed and analyzed extracellular single-neuron recordings from the medial superior temporal area (MST) of two rhesus monkeys performing a spatial and feature-based attention task. Both shifting spatial attention into the RF and deploying feature-based attention to the preferred direction (relative to the non-preferred direction) enhanced the firing rate of MST neurons, as expected based on previous studies (Treue and Maunsell 1996; Treue and Martinez Trujillo 1999; Patzwahl and Treue 2009). In addition, spatial attention also led to a concurrent net reduction in burstiness, as reported earlier from V4. However, feature-based attention is not associated with a significant reduction in burstiness, though it did enhance firing rate. This absence of significant burstiness reduction cannot be explained by the smaller effect size of feature-based attention compared with spatial attention. Furthermore, the effects of spatial attention on firing rate and burstiness could be dissociated. Our results extend our understanding of the attentional effects on the temporal patterns of action potential discharge and support the idea that different types of attention may involve different physiological mechanisms.

Materials and Methods

Animal Use and Surgical Procedures

Data were collected from two male rhesus monkeys (*Macaca mulatta*, Monkey W, Monkey N, both 12-year-old males). Area MST was accessible through a recording chamber implanted over the parietal lobe based on a magnetic resonance imaging (MRI) scan (right hemisphere for Monkey W, left hemisphere for Monkey N). Each monkey was implanted with a titanium

head holder to minimize head movements during the experiment. Both monkeys were seated in custom-made primate chairs and head-fixed during the experiment. All procedures were conducted in accordance with German laws governing animal care and approved by the district government of Oldenburg, Lower Saxony, Germany. Surgeries were conducted under general anesthesia and post-surgical care using standard techniques.

Experimental Setup

The monkeys performed the tasks in a dimly lit room, with the only source of light being the display monitor. A custom computer program for experiment control, running on an Apple Macintosh PowerPC handled the stimulus presentation, eye-position control, as well as data collection and storage. Eye positions were monitored with a video-based eye tracker (ET49, sampling rate 60 Hz; Thomas Recording, Giessen, Germany). A CRT monitor placed at a distance of 57 cm from the monkey was used to display the visual stimulus at a refresh rate of 60 Hz and a spatial resolution of 40 pixels per degree. The monitor covered approximately $40^\circ \times 30^\circ$ of visual angle.

Electrophysiological Procedures

We recorded neuronal activity extracellularly using a three-channel microdrive system (Mini Matrix; Thomas Recording) and a Multichannel Acquisition Processor system (Plexon, Inc., Dallas, TX), running at a sampling rate of 40 kHz. Action potentials were sorted online (waveform window discrimination, Sort Client; Plexon Inc.) and recorded. MST was identified by referencing the recordings to the structural MRI and by the physiological properties of the recorded neurons (large RFs compared with MT and direction tuning to spiral motion; Graziano et al. 1994). We recorded data from well-isolated neurons if their response to the preferred spiral motion direction was at least twice as high as the response to the null direction. Six recorded neurons were excluded from this population as we were unable to record at least three hit trials for each trial condition. Once a neuron was isolated, its RF was estimated by manually moving a static stimulus on the monitor while the monkey maintained his gaze on the fixation task. Once the RF was identified, a series of spiral motion stimuli were presented in the RF in sequence in order to determine the feature preference of the neuron. We used 12 spiral motion directions. The direction that elicited the highest response was taken as the "preferred direction" of the unit, while the opposite direction was taken as the "null direction." After this phase of initial characterization, the monkeys performed different experimental tasks while the neuron's activity was recorded.

Behavioral Task

We analyzed three different conditions from the cued detection task in this study. In cued detection trials, the monkeys had to respond to a speed change in 1 of 2 spiral motion stimuli (the target, identified by a preceding stationary cue presented at the same location) while ignoring similar changes in the other stimulus (the distractor). The spiral motion stimuli were random dot patterns (RDPs) in which the motion direction of all dots in a given RDP maintains a constant angle with the radial axis (Fig. 1A). MST neurons are known to be tuned for this "spiral direction" (Graziano et al. 1994). The RDPs had a diameter of 4° of visual angle and a dot density of 8 per square degree. The

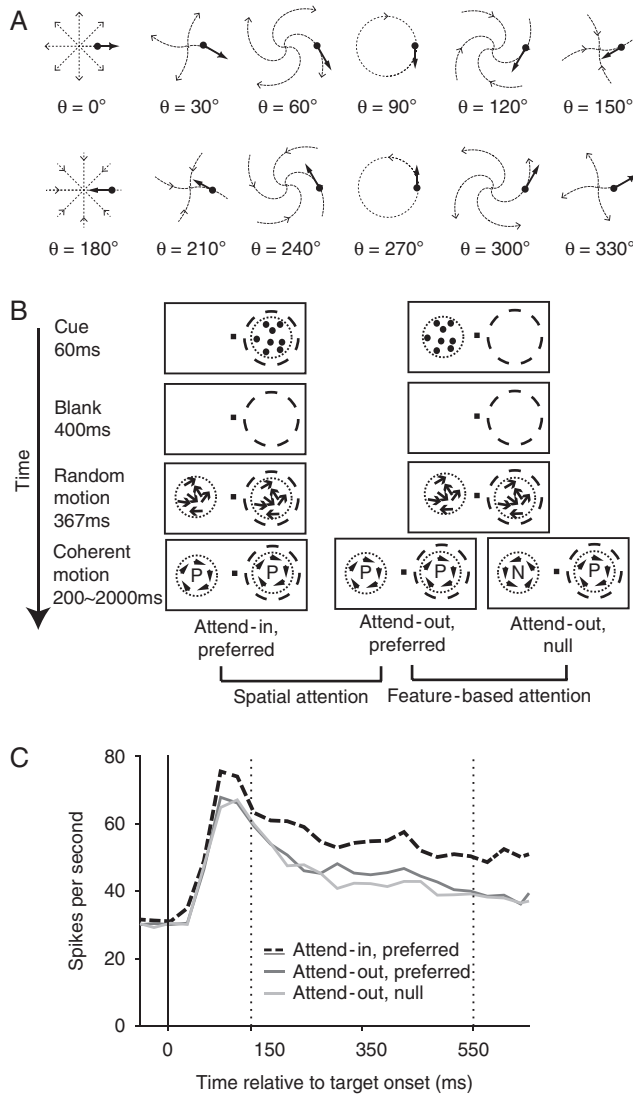


Figure 1. Behavioral task and neuronal responses. (A) Spiral motion stimuli. In each RDP stimulus, all dots move in directions that maintain a constant angle θ with the radial axis (see the thick black arrows), which defines the “spiral direction” of the RDP. The direction that elicited the highest response in a given neuron was taken as the “preferred direction” of the unit, while the opposite direction was taken as the “null direction.” (B) Trial sequence. Once the monkey depressed a lever and foveated the central fixation point (black square), a spatial cue (stationary RDP) briefly appeared either in or outside the RF of the recorded neuron (dashed circle). After a blank interval, two RDPs in non-coherent motion were presented. After 367 ms both stimuli became fully coherent, preferred (clockwise rotation in this example) or null (anti-clockwise rotation in this example) direction motion patterns. The monkey had to respond within 400 ms to a speed change in the cued stimulus (the “target”) to correctly complete the trial. The horizontal brackets indicate the conditions that were compared to establish the modulation by spatial or feature-based attention, respectively. (C) Average PSTH across our population of 100 MST neurons, with time relative to target and distractor onset (solid vertical line) in all three attentional conditions. Vertical dotted lines indicate the start and end of the analysis period.

luminance of the dots was 75 cd/m^2 on a gray background of 35 cd/m^2 .

The monkey started each trial by touching a lever and directing its gaze onto a central fixation point ($0.2^\circ \times 0.2^\circ$). Throughout the trial, the monkeys were required to maintain their gaze within 1.8 degrees of the fixation point, or the trial was aborted. After 150 ms from the start of the trial, a static

RDP was presented as a spatial cue for 67ms. After cue offset, a 400 ms blank period followed. The blank period ended with the onset of two zero-coherence spiral motion RDPs: one at the cued target location and the other at a location symmetrically opposite to it (i.e. reflected around the fixation point). After 367 ms, both RDPs turned into fully coherent spiral motions; this time point was defined as the target onset. The monkey had to respond within 400 ms (by releasing the lever) to a change in speed of the target RDP. At the same time, the monkey had to ignore any changes in the distractor, that is, the RDP at the uncued location. Each correctly completed trial was rewarded with juice. The target speed change time was randomly chosen for each trial from between 250 and 2500 ms after target onset.

For this study, we analyzed three behavioral conditions adapted from Treue & Martinez-Trujillo (1999) to determine the effects of spatial and feature-based attention: the *attend-in preferred* condition, the *attend-out preferred* condition, and the *attend-out null* condition. In all three conditions, the RDP in the RF moved in the preferred direction of the neuron. In the attend-in preferred condition, the RDP inside the RF was the target and the distractor RDP (outside the RF) also moved in the preferred direction. In the attend-out preferred and the attend-out null conditions, the RDP outside the RF was the target and moved in the preferred direction for the attend-out preferred condition and in the null direction for the attend-out null condition. Comparing neuronal responses in the attend-in preferred and attend-out preferred conditions isolates the effects of spatial attention, while comparing attend-out preferred to attend-out null isolates the effects of feature-based attention. Trials from the three conditions were performed in an interleaved manner.

Data Analysis

All data analysis was performed using custom software in MATLAB R2015a (MATLAB Inc., Natick, MA). We included data from all neurons that showed a tuning for spiral motion direction, with the preferred direction position-invariant, that is, unaffected by placing the spiral motion at different positions within the RF (Graziano et al. 1994). We only included correctly performed trials in our analysis. Peri-stimulus time histograms (PSTHs) in Figure 1C were calculated using non-overlapping 30 ms bins. The mean activity for each neuron across trials was first calculated and then these mean PSTHs for individual neurons were averaged across neurons to obtain the displayed PSTHs.

Burst Analysis

Burstiness was estimated for each neuron and each task condition, during an analysis period from 150 to 550 ms after target (and distractor) onset. We picked 150 ms as the start of the analysis window to exclude the transient activity induced by the coherent motion onset, and 550 ms as the end of the analysis window to ensure enough trials for the burstiness calculation where no motion change occurred in either the target or distractor RDP within the analysis window. We selected for analysis all correctly completed trials with neither a distractor speed change nor a target speed change during the analysis period. Only neurons with at least three such trials for each attentional condition were included. To quantify burstiness, we used the same approach described in 2 earlier studies (Compte et al. 2003; Anderson et al. 2011). For a set of trials from each

neuron and each attentional condition, we first calculated the mean autocorrelation function (ACF). We then calculated a shuffle predictor defined as the cross-correlation function across all pairs of trials in the set. To obtain a normalized ACF, we subtracted the mean of the shuffle predictor from the mean ACF and normalized the difference with the standard deviation of the shuffle predictor. Burstiness was then defined as the average height of the normalized ACF for time lags from 1 to 4 ms. This burstiness measure is also partially similar to that in [Anderson et al. \(2013\)](#), with the difference that [Anderson et al. \(2013\)](#) normalized by the mean shuffle predictor (rather than its standard deviation) and further multiplied the normalized ACF by the impulse response of a band-pass filter (10–40 Hz) and integrated the result to obtain a burstiness value. The [Anderson et al. \(2013\)](#) procedure calculates burstiness by computing a weighted sum of the normalized ACF between 1 and 11 ms, 33 and 46 ms, 55 and 77 ms, and so on until 256 ms, and subtracts a weighted sum of the ACF between 11 and 33 ms, 46 and 55 ms, 77 and 92 ms, and so on until 256 ms. Since this decaying and roughly sinusoidal weighting function has a band-pass frequency spectrum from 10 to 40 Hz, this also has the effect of integrating the Fourier transform of the ACF between 10 and 40 Hz. Though this procedure appears quite different from the [Anderson et al. \(2011\)](#) procedure that we use, the burstiness values it generates are highly correlated with the ones generated by the [Anderson et al. \(2011\)](#) procedure, and our interpretations and conclusions remain the same using either measure (see Results).

Quantifying Attentional Modulation

We quantified the magnitude of attentional modulation of firing rate using a very common attentional index, defined as the difference of values between attentional conditions normalized by their sum. Specifically, the attentional index of spatial attention on firing rate (denoted by FR) was calculated as:

$$AI_{FR}^{spatial} = \frac{FR_{in} - FR_{out}}{FR_{in} + FR_{out}}$$

where in and out refer to the conditions with spatial attention into and outside the RF (with both RDPs always moving in the preferred direction).

Similarly, the attentional index of feature-based attention on firing rate was calculated as:

$$AI_{FR}^{feature} = \frac{FR_{pref} - FR_{null}}{FR_{pref} + FR_{null}}$$

where pref and null refer to attention to the preferred or null direction RDP outside the RF (with the preferred direction distractor RDP inside the RF).

Unlike firing rate, burstiness values using our measure can have values below 0, and the attentional index as defined above only works for non-negative values. We therefore simply use the difference of burstiness values between attentional conditions to quantify the attentional effect on burstiness. Finally, for both firing rate and burstiness, we report the averages using medians (after converting the median attentional index back to a percentage value) and use the Wilcoxon signed-rank test to assess statistical significance. We use the Kendall rank correlation to measure potential associations and determine statistical significance.

Trial-Swap Analysis

To determine whether changes in firing rate could be dissociated from changes in burstiness, we also performed an analysis within individual neurons where for each neuron, we created new data sets by exchanging trials with similar firing rate between the attend-in preferred and attend-out preferred conditions. The goal of this trial-swap was to exchange as many trials as possible (with similar firing rates) between the attend-in and attend-out conditions, so that the mean firing rates of the 2 conditions only changed minimally, and even this minimal change always led to a greater enhancement of firing rate by spatial attention. Specifically, we first sorted the trials in the attend-in and attend-out conditions by their spike count. To choose trials for swapping, we created 2 subsets of trials: for Subset 1, we picked the N least spiking trials from the condition with higher mean firing rate, and for Subset 2, we picked the N most spiking trials from the condition with lower firing rate. N was chosen to be the largest number that ensured that the subset from the attend-in condition has lower mean firing rate than the subset from the attend-out condition. Chosen in this manner, swapping Subset 1 with Subset 2 retained or enhanced the attentional index for each neuron (Fig. 3C) and therefore predicted a larger reduction for burstiness as well, if the attentional effect on burstiness was coupled to that on firing rate.

Waveform Duration Calculation

We recorded each spike waveform over a 800ms window with a sampling rate of 40 kHz. For each cell, we first normalized the height of the spike waveforms by calculating its z-score relative to the average of the waveform over time. The normalized waveforms were then aligned to the trough of each waveform, averaged to obtain a mean waveform and then interpolated to a time resolution of 1μs (using the MATLAB “interp1” function, “spline” mode). Waveform duration was then defined as the time duration between the trough and the following peak of the interpolated mean waveform.

Results

We investigated the burstiness of 100 MST neurons from 2 monkeys (44 neurons for Monkey N, 56 neurons for Monkey W) in 3 attentional conditions (Fig. 1B), in which the monkeys were cued to attend to the spatial location and motion direction of a target stimulus in the presence of a second, distractor stimulus. In all 3 conditions, the physical stimulus within the neuron's RF was a RDP moving in the preferred direction. Monkey N correctly responded to the target change 92.8% of the time (hit rate), and missed the remaining 7.2 % of changes; releases before the target change occurred on 6.8% of trials (early release rate). Monkey W had a hit rate of 93.4% and an early release rate of 6.5%. The mean reaction times were 333 ms (standard deviation: 41 ms) for Monkey N and 319 ms (SD = 56ms) for Monkey W.

Both Spatial and feature-based attention Modulate Firing Rate

When both RDPs moved in the preferred direction of the recorded neuron, comparing the responses when the RDP in the RF was the target to that when it was the distractor enabled us to examine the effects of spatial attention. Similarly, when the distractor was in the RF (and moved in the preferred

direction), comparing the responses when the target RDP outside the RF moved in the preferred direction to that when it moved in the anti-preferred direction enabled us to examine the effects of feature-based attention. The average population PSTH (Fig. 1C) shows a clear enhancement of the firing rate by both spatial attention (slashed black curve compared with solid black curve) and feature-based attention (solid black curve compared with solid gray curve). Similarly, the attentional index of firing rate (Fig. 2B) shows a clear enhancement by both spatial and feature-based attention: the median increase in firing rate for spatial attention is 19.4% ($P < 0.0001$) in the overall

population (Monkey N: 16.0%, $P < 0.0001$, monkey W: 21.8%, $P < 0.0001$) and the median increase in firing rate for feature-based attention is 6.7% ($P < 0.0001$) in the overall population (Monkey N: 6.6%, $P < 0.0001$, Monkey W: 6.7%, $P = 0.005$). The effect of spatial attention on firing rate is significantly greater than that for feature-based attention (overall population: $P = 0.0002$, 0.0001, Monkey N: $P = 0.04$, Monkey W: $P = 0.002$).

Spatial Attention Significantly Reduces Burstiness; feature-based attention Does not

Spatial attention clearly reduces burstiness in our population of MST neurons. As depicted for an example neuron (Fig. 2A), the normalized ACF shows a clear reduction in height when spatial attention was directed into the RF, compared with when it was directed outside the RF (black dashed curve compared with grey curve). In the population, spatial attention led to a median reduction in burstiness of 0.105 in the overall population ($P = 0.0001$; Monkey N: median = -0.091 , $P = 0.006$; Monkey W: median = -0.114 , $P = 0.005$). However, feature-based attention did not significantly reduce burstiness in either of the 2 monkeys. It led to an increase in burstiness in Monkey W (median = 0.095 , $P = 0.04$), and no significant effect on burstiness in Monkey N (median = -0.117 , $P = 0.1$); this difference between the monkeys was statistically significant (rank sum test, $P = 0.02$). This was true even if the analysis window's beginning was shifted to 300 ms following RDP onset, in order to account for the delayed emergence of feature-based attention (black vs. gray curves in Fig. 1C): Again, Monkey W showed a significant increase (median = 0.082 , $P = 0.04$) and Monkey N showed no significant effect (median = -0.055 , $P = 0.2$). A direct pairwise comparison showed that spatial attention led to a larger reduction in burstiness than feature-based attention in both monkeys, but the effect was not significant in Monkey N (median additional effect with spatial attention in Monkey W = -0.136 , $P = 0.008$; in Monkey N = -0.062 , $P = 0.6$).

These conclusions do not depend on our specific implementation of the burstiness measure: using the burstiness measure from Anderson, Mitchell and Reynolds (2013), spatial attention again reduced burstiness (overall population: median reduction = 0.042 , $P < 0.0001$; Monkey N: median reduction = 0.025 , $P = 0.007$; Monkey W: median reduction = 0.066 , $P = 0.0006$), and there was no significant effect on burstiness with feature-based attention (overall population: median increase = 0.002 , $P = 0.7$; Monkey N: median increase = 0.001 , $P = 0.8$; Monkey W: median increase = 0.023 , $P = 0.5$).

Burstiness and Firing Rate Modulation by Spatial Attention Are Uncorrelated

A recent computational model proposed that firing rate increases and burstiness reductions by spatial attention emerge via a common mechanism. A firing rate increase may also lead to a reduction in our burstiness measure in the presence of a refractory period (see Discussion). However, we find that the burstiness reduction by spatial attention can be dissociated from the concurrent firing rate increase. The effects of spatial attention on firing rate and burstiness were not correlated in the population of recorded neurons: the correlation coefficient between the change of firing rate and change of burstiness with spatial attention was -0.081 ($P = 0.2$) in the overall population, -0.042 ($P = 0.7$) in Monkey N, and -0.119 ($P = 0.2$) in Monkey W. We also performed an analysis within individual neurons, where for each neuron, we created new data sets by

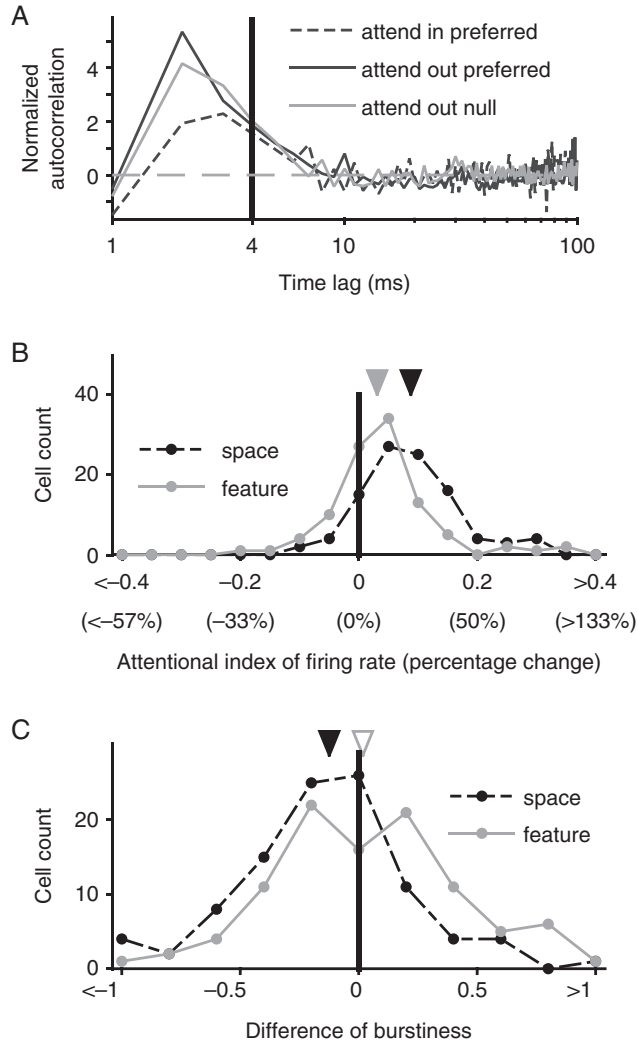


Figure 2. Spatial and feature-based attentional modulation of firing rate and burstiness. (A) The normalized autocorrelation of an example MST neuron in different attentional conditions. The y-axis and the vertical solid line at 4 ms time lag demarcate the range of time lags used to estimate the burstiness index. Burstiness is defined as the average height of the normalized autocorrelation for the time lags between 1 and 4 ms. The horizontal dashed line indicates the normalized autocorrelation for a Poisson spike train. (B) Distribution of attentional indices of firing rates with spatial attention (black dashed line) and feature-based attention (gray solid line). (C) Distribution of attentional indices of burstiness for spatial attention (black dashed line) and feature-based attention (gray solid line). In B and C, black and gray triangles indicate median index values for spatial and feature-based attention, respectively. Filled triangles indicate significant deviations from zero (signed-rank test, see Materials and Methods), open triangles indicate lack of significance. X-axis values in brackets represent percentage changes: positive values are increases and negative values are reductions.

exchanging trials with similar firing rate between the attend-in and attend-out conditions (see Materials and Methods for detailed algorithm). The goal of this trial-swap was to exchange as many trials as possible (with similar firing rates) between the attend-in and attend-out conditions, so that the mean firing rates of the two conditions only changed minimally, and even this minimal change always led to a greater enhancement of firing rate by spatial attention (Fig. 3C). We reasoned that if

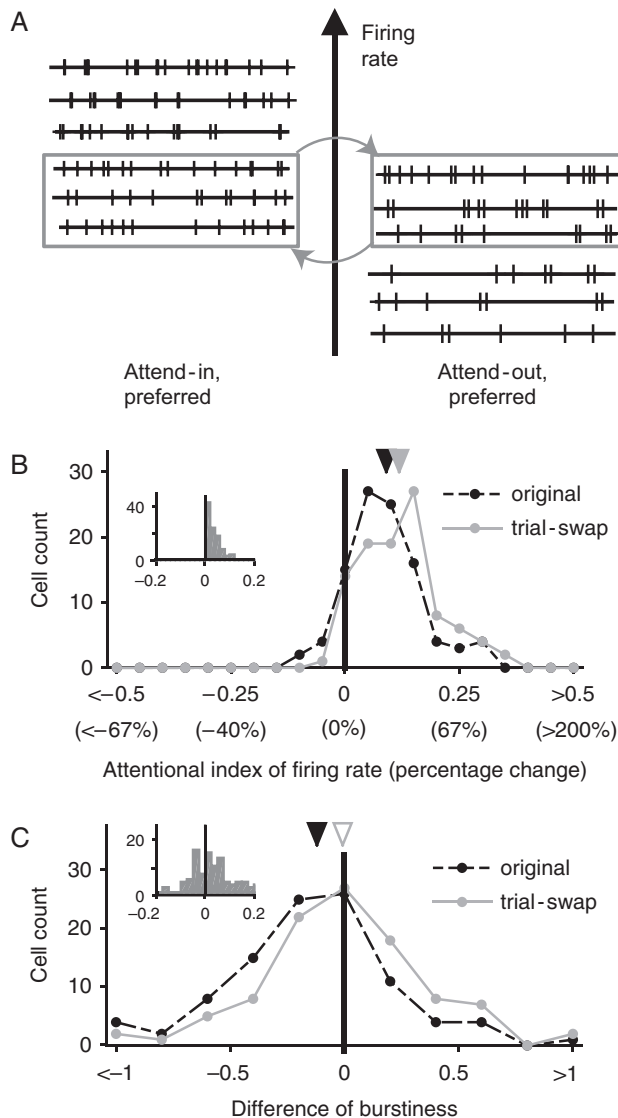


Figure 3. The reduction in burstiness by spatial attention can be dissociated from the firing rate increase. (A) Illustration of the trial-swap process. For each neuron, trial subsets with similar mean firing rates were swapped between spatial attention conditions. See Materials and Methods for details. (B) Distribution of attentional indices of firing rate across the MST population with spatial attention before (black dashed line) and after (gray solid line) swapping trials with similar firing rate. The inset shows the distribution of pairwise differences between attentional indices before and after the trial-swap. Trial-swaps were done conservatively, such that any change in the attentional index of firing rate was an increase. (C) Distribution of attentional indices of burstiness with spatial attention before (black dashed line) and after (gray solid line) swapping trials. Black and gray triangles indicate median index values for spatial and feature-based attention, respectively. Filled triangles indicate significant deviations from zero (signed-rank test, see Materials and Methods), open triangles indicate lack of significance. The inset shows the distribution of pairwise differences between attentional indices after and before trial-swap. The burstiness reduction effect disappeared after the trial-swap.

the burstiness effect of attention was linked to the firing rate, then exchanging trials with similar firing rate between the two attentional conditions in this manner would either slightly enhance the attentional reduction of burstiness or leave it unchanged. However, if the burstiness reduction in the attend-in condition was linked to the attentional state (and not the firing rate), then exchanging trials between the attend-in and attend-out conditions would substantially reduce the burstiness reduction by attention. Our results support this latter hypothesis. As expected from the choice of trials, trial-swap slightly increased the median attentional index of firing rate: the median increase in the attentional index of firing rate after the swap was 0.023 in the overall population, 0.017 in Monkey N, and 0.034 in Monkey W, and every neuron showed a slight increase. However, the reduction of burstiness by spatial attention was no longer statistically significant after the trial-swap: the median difference in burstiness between the attend-in preferred and attend-out preferred conditions after the trial-swap was -0.008 ($P = 0.8$) in the overall population, -0.008 ($P = 0.3$) in Monkey N, and 0.043 ($P = 0.5$) in Monkey W (the values before the trials swap were -0.105 in the overall population, -0.091 in Monkey N, and -0.114 in Monkey W, as reported above). A pairwise comparison within neurons also indicated that there was a reduction in the effect of spatial attention on burstiness as a result of the trial-swap, though this reduction was only statistically significant in one monkey (median reduction in the burstiness differences between attend-in preferred and attend-out preferred conditions due to the trial-swap was 0.091, $P = 0.03$ in the overall population, 0.045, $P = 0.6$ in Monkey N, and 0.132, $P = 0.02$ in Monkey W). Overall, these results indicate that the burstiness reduction by spatial attention can be dissociated from its effects on firing rate.

Differences in Burstiness Modulation by Spatial and Feature-based Attention Are not Confounded by Attentional Effect Size

Both spatial and feature-based attention are associated with increases in the firing rate of MST neurons, in line with previous reports. We showed above that spatial attention was associated with a significant reduction of burstiness, while feature-based attention did not significantly reduce burstiness: feature-based attention actually increased burstiness significantly in Monkey W and led to a non-significant effect in Monkey N. However, one could argue that the lack of a significant burstiness reduction with feature-based attention may simply be a result of the smaller effect of feature-based attention (evidenced by a 6.7% median increase in firing rate) compared with spatial attention (with a 19.4% increase). We therefore examined the relationship between the attentional effects on firing rate and burstiness in the spatial and feature-based attention conditions (Fig. 4). There was no significant correlation between the attentional effects on firing rate and burstiness in either condition (spatial attention: $\tau = -0.031$ ($P = 0.6$) in the overall population, $\tau = -0.011$ ($P = 0.9$) in Monkey N, and $\tau = -0.068$ ($P = 0.5$) in Monkey W; feature-based attention: $\tau = 0.021$ ($P = 0.8$) in the overall population, $\tau = -0.011$ ($P = 0.9$) in Monkey N, and $\tau = -0.068$ ($P = 0.5$) in Monkey W), indicating that the reduction of burstiness by spatial attention could not be explained by its larger effect on firing rate. Similarly, no significant correlation was found when using the difference of firing rates as the measure instead of the attentional index (all P -values ≥ 0.5). Furthermore, we examined the effect of spatial attention on burstiness in neurons where the spatial

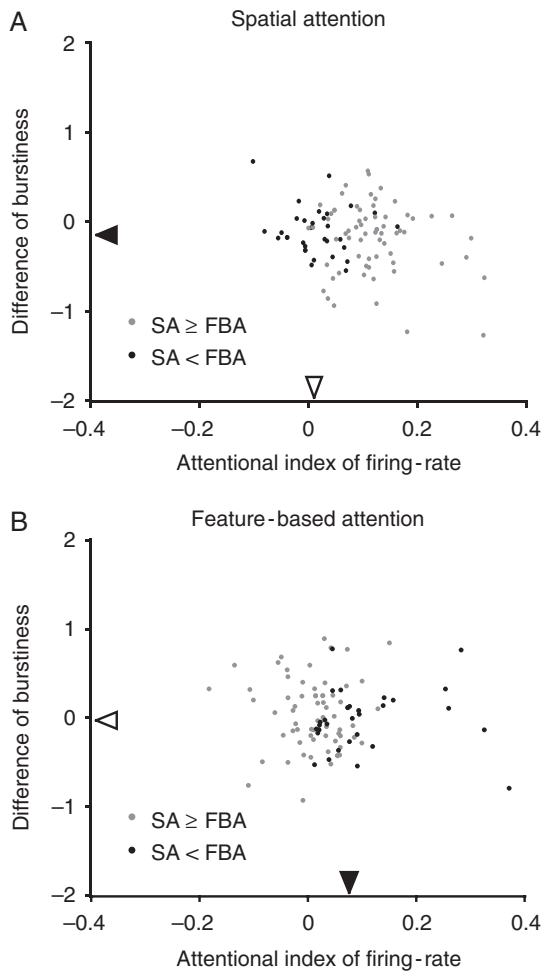


Figure 4. The lack of significant burstiness reduction by feature-based attention cannot be explained by the smaller effect size of FBA. (A) The effects of spatial attention on firing rate (abscissa) and burstiness (ordinate). Black dots represent neurons with a larger firing rate effect of feature-based attention ('FBA' in the figure) than of spatial attention ('SA' in the figure), while gray dots represent the remaining neurons. Both subpopulations (black and gray dots) showed a significant burstiness reduction but no correlation between firing rate and burstiness modulation (see text). In the figure, SA refers to spatial attention and FBA refers to feature based attention. (B) The effects of feature-based attention on firing rate and burstiness (plotting conventions as in A). In both panels, the black triangles indicate median index values for black dots on the respective axis; Filled triangles indicate significant deviations from zero (signed-rank test, see Materials and Methods), open triangles indicate lack of significance. Neither subpopulation (black and gray dots) showed a significant burstiness reduction or a correlation between firing rate and burstiness modulation (see text).

attentional indices of firing rate were smaller than the median feature-based attention index of firing rate (which equals 0.0323, and is also denoted by the filled gray triangle in Fig. 2A; the median index is 0.0321 in Monkey N, 0.0323 in Monkey W). Even within this subset of cells, spatial attention was associated with a reduction in burstiness (two monkeys pooled: median = -0.104 , $P = 0.02$, $n = 24$; Monkey N: median = -0.154 , $P = 0.06$, $n = 10$; Monkey W: median = -0.087 , $P = 0.03$, $n = 14$). Consistent with the lack of correlation between the attentional effects on firing rate and burstiness, this reduction in burstiness was not significantly different from that of the remaining neurons (rank sum test, $P = 0.8$ in overall population, $P = 0.8$ in Monkey N, and $P = 0.6$ in Monkey W). Along similar lines, we also looked at another subset of neurons whose spatial

attentional index for firing rate was actually less than that for feature-based attention (black dots in Fig. 4). Again, even in this subset of neurons where the measured spatial attentional effect on firing rate was less than that of feature-based attention, spatial attention was still associated with a significant reduction in burstiness (two monkeys pooled: median = -0.133 , $P = 0.01$, $n = 32$, filled black triangle on vertical axis in Fig. 4A; Monkey N: median = -0.080 , $P = 0.1$, $n = 16$, Monkey W: median = -0.157 , $P = 0.03$, $n = 16$). This reduction in burstiness by spatial attention in this subset was not significantly different from that in the remaining neurons (for both monkeys pooled: median = -0.098 , $P = 0.002$, rank sum test $P = 0.8$; Monkey N: median = -0.091 , $P = 0.01$, rank sum test $P = 0.7$; Monkey W: median = -0.105 , $P = 0.03$, rank sum test $P = 0.98$). These results suggest that the significant reduction of burstiness associated with spatial attention (compared with feature-based attention) cannot be explained on the basis of a difference in effect-sizes (as measured by the effects of the two attention types on firing rate).

Effects of Spike Waveform Duration

Anderson et al. (2013) reported that V4 neurons could be categorized into 2 classes based on a bimodal population distribution of waveform durations, and burstiness effects were only found in the broad-spiking population. Our neuronal population did not show a bimodality in waveform durations (Hartigan's dip test: Monkey N, $P = 0.4$ and Monkey W, $P = 1$; Supplementary Figure 2).

Discussion

Burstiness, defined as the tendency of a neuron to discharge discrete groups of consecutive action potentials, has been extensively identified from both in vitro and in vivo recordings in various neuron types and brain regions (McCormick et al. 1985; Bair et al. 1994; Compte et al. 2003). Our data demonstrate that spatial attention is associated with a reduction of burstiness in MST neurons. We disambiguate this reduction of burstiness from the concurrent increase in firing rate induced by spatial attention. These results suggest that burstiness reduction might be a ubiquitous effect of spatial attention across visual areas.

The functional properties and neural utility of bursts in spike trains have been a topic of much speculation and interest over the years (Cattaneo et al. 1981; Izhikevich et al. 2003; Krahe and Gabbiani 2004; Shih et al. 2011). Since a burst of spikes may induce a stronger change in postsynaptic potential than more temporally dispersed spikes, it has been suggested that bursts are a more reliable unit to transmit information (Izhikevich 2007). In line with this assertion, other studies have also suggested that bursts enhance functional connectivity between areas (Bonifazi et al. 2009; Kwan and Dan 2012; Womelsdorf et al. 2014). The prevailing hypothesis that bursts enhance information encoding transfer predicts that burstiness would increase when attention is directed towards the RF, so that attended neural responses have an advantage in transmission (Anderson et al. 2013). However, the data indicate the opposite. One possible resolution is that the multiple spikes in a burst carry redundant sensory information: it has been shown that the event rate, where each event is either a single burst of spikes or an isolated spike, is on average a more sensitive measure of direction selectivity than the total number of spikes (Bair et al. 1994). Spatial attention would then make more efficient use of each spike by reducing this redundancy.

A previous study from V4, in the ventral cortical visual processing stream, also reported a reduction of burstiness in the broad-spiking neurons of V4 when spatial attention is directed into their RFs; their median effect size is less than the one we report here for MST (Anderson et al. 2013). One other previous study from V4 (McAdams and Maunsell 1999) also looked at the effect of spatial attention on burstiness. McAdams and Maunsell reported that they found an increase in the rate of bursting with spatial attention, but that this could be accounted for by the higher firing rate with spatial attention. As also pointed out by Anderson et al. (2013), this seemingly contradictory finding was based on a measure of burstiness that increased with firing rate, and the across-neuron analysis used in that study was also less sensitive than the within-neuron analysis that we and Anderson et al. (2013) use. It appears probable that this explains the discrepancy between the findings of McAdams and Maunsell and those of our study as well as Anderson et al. (2013).

Based on their results from area V4, Anderson et al. (2013) proposed a model that accounts for the modulation of firing rate and burstiness by spatial attention via a common cellular mechanism that increases both inhibitory and excitatory synaptic conductances. Our data instead suggest that the effects of spatial attention on burstiness and firing rate stem, at least in part, from separate mechanisms, since in MST the spatial attentional modulation of burstiness and firing rate can be dissociated. This dissociation also argues against another potential explanation for the reduction of burstiness by spatial attention: in a spike train with a refractory period (of say 4 ms), any increase in firing rate will leave the autocorrelation function between time lags of 1 and 4 ms unaffected but will increase the magnitude of the cross-correlation function (shuffle predictor) and therefore, our burstiness measure will decrease with firing rate under these conditions. However, this does not appear to be the case with our results.

In our experiment, monkeys had to maintain fixation within 1.8 degrees of the central fixation point. It is possible that the monkeys made small eye movements within this window during fixation, and that the neural effects of these eye movements (Bair and O'Keefe 1998; Martinez-Conde et al. 2002; Hafed and Krauzlis 2010) interact with the effects of attention to differentially modulate burstiness in the attentional conditions we test. However, due to the low sampling rate of the eye tracker we used (60 Hz), we are unable to perform a reliable analysis of microsaccade effects. Evaluating the role of microsaccades in the burstiness reduction with spatial attention remains a topic for future studies.

Our findings have implications in the context of the feature similarity gain model of visual attention (Treue and Maunsell 1999; Martinez-Trujillo and Treue 2004; Maunsell and Treue 2006), which proposes that the gain of a visual neuron is maximal when the attended feature (or spatial location) matches the neuron's preferred feature (or spatial location). The model supports (but does not require) a unified mechanism for spatial and feature-based attention. In other words, spatial location is just another feature. Our observation of different effects of spatial and feature-based attention on burstiness suggests at least 2 possibilities. Spatial and feature-based attention mechanisms may differ but happen to have qualitatively similar effects on firing rate. Alternatively, both types of attention may share a neural mechanism that modulates firing rate but spatial attention additionally engages a separate mechanism that affects burstiness. Our evidence for, at least partially, separate mechanisms of spatial and feature-based attention is

supported by psychophysical studies that have observed differences in the effects of spatial and feature-based attention on human visual perception (Ling et al. 2009) and by recordings from V4 neurons that have observed differences in firing rate modulation by spatial and feature-based attention (Hayden and Gallant 2005; David et al. 2008; Cohen and Maunsell 2011). Furthermore, while the impact of the burstiness reduction by spatial attention on the functional and behavioral consequences of spatial attention remains unknown, our data suggest that the consequences of feature-based attention do not depend on burstiness modulation.

Overall, our results from the dorsal visual pathway suggest that the modulation of burstiness by spatial attention is a general phenomenon across visual cortex that arises from a common neural mechanism. Spatial and feature-based attention, however, may involve partially different underlying neural mechanisms. Further studies of these differences may yield a fuller understanding of the common and unique aspects of various types of attention across visual cortex and how they reflect the underlying attentional neural mechanisms.

Author Contribution

S.B.R., D.K., and S.T. designed the experiment; S.B.R. and D.K. performed the experiment; C.X. and B.S.K. designed the analysis; C.X. and B.S.K. analyzed the data; C.X., B.S.K. and S.T. wrote the paper

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Funding

Supported by the Deutsche Forschungsgemeinschaft (Collaborative Research Center 889 "Cellular Mechanisms of Sensory Processing" and the Research Unit 1847 "Physiology of Distributed Computing Underlying Higher Brain Functions in Non-Human Primates").

Notes

We thank Emily Anderson and Jude Mitchell for sharing their implementation of the burstiness measure in (Anderson et al. 2013). *Conflict of Interest:* None declared.

References

- Anderson EB, Mitchell JF, Reynolds JH. 2011. Attentional modulation of firing rate varies with burstiness across putative pyramidal neurons in macaque visual area V4. *J Neurosci.* 31:10983–10992.
- Anderson EB, Mitchell JF, Reynolds JH. 2013. Attention-dependent reductions in burstiness and action-potential height in macaque area V4. *Nat Neurosci.* 16:1125–1131.
- Bair W, Koch C, Newsome W, Britten K. 1994. Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J Neurosci.* 14:2870–2892.
- Bair W, O'Keefe LP. 1998. The influence of fixational eye movements on the response of neurons in area MT of the macaque. *Vis Neurosci.* 15:779–786.
- Bichot NP, Heard MT, DeGennaro EM, Desimone R. 2015. A source for feature-based attention in the prefrontal cortex. *Neuron.* 88:832–844.

- Bisley JW. 2011. The neural basis of visual attention. *J Physiol.* 589:49–57.
- Bonifazi P, Goldin M, Picardo MA, Jorquera I, Cattani A, Bianconi G, Represa A, Ben-Ari Y, Cossart R. 2009. GABAergic hub neurons orchestrate synchrony in developing hippocampal networks. *Science.* 326:1419–1424.
- Carrasco M. 2011. Visual attention: the past 25 years. *Vision Res.* 51:1484–1525.
- Cattaneo A, Maffei L, Morrone C. 1981. Patterns in the discharge of simple and complex visual cortical cells. *Proc R Soc Lond B Biol Sci.* 212:279–297.
- Cohen MR, Kohn A. 2011. Measuring and interpreting neuronal correlations. *Nat Neurosci.* 14:811–819.
- 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.
- Compte A, Constantinidis C, Tegner J, Raghavachari S, Chafee MV, Goldman-Rakic PS, Wang XJ. 2003. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J Neurophysiol.* 90:3441–3454.
- David SV, Hayden BY, Mazer JA, Gallant JL. 2008. Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron.* 59:509–521.
- Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu Rev Neurosci.* 18:193–222.
- Esguaje M, Daliri MR, Treue S. 2015. Attention decreases phase-amplitude coupling, enhancing stimulus discriminability in cortical area MT. *Front Neural Circuits.* 9:82.
- Fries P. 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu Rev Neurosci.* 32:209–224.
- Graziano MS, Andersen RA, Snowden RJ. 1994. Tuning of MST neurons to spiral motions. *J Neurosci.* 14:54–67.
- Hafed ZM, Krauzlis RJ. 2010. Microsaccadic suppression of visual bursts in the primate superior colliculus. *J Neurosci.* 30:9542–9547.
- Hayden BY, Gallant JL. 2005. Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4. *Neuron.* 47:637–643.
- Izhikevich EM. 2007. *Dynamical systems in neuroscience.* London: The MIT Press.
- Izhikevich EM, Desai NS, Walcott EC, Hoppensteadt FC. 2003. Bursts as a unit of neural information: selective communication via resonance. *Trends Neurosci.* 26:161–167.
- Krahe R, Gabbiani F. 2004. Burst firing in sensory systems. *Nat Rev Neurosci.* 5:13–23.
- Kwan AC, Dan Y. 2012. Dissection of cortical microcircuits by single-neuron stimulation in vivo. *Curr Biol.* 22:1459–1467.
- Ling S, Liu T, Carrasco M. 2009. How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Res.* 49:1194–1204.
- Lisman JE. 1997. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci.* 20:38–43.
- Martinez-Conde S, Macknik SL, Hubel DH. 2002. The function of bursts of spikes during visual fixation in the awake primate lateral geniculate nucleus and primary visual cortex. *Proc Natl Acad Sci U S A.* 99:13920–13925.
- 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, Treue S. 2006. Feature-based attention in visual cortex. *Trends Neurosci.* 29:317–322.
- McAdams CJ, Maunsell JH. 1999. Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron.* 23:765–773.
- McCormick DA, Connors BW, Lighthall JW, Prince DA. 1985. Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *J Neurophysiol.* 54:782–806.
- Mitchell JF, Sundberg KA, Reynolds JH. 2009. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron.* 63:879–888.
- Moore T, Armstrong KM. 2003. Selective gating of visual signals by microstimulation of frontal cortex. *Nature.* 421:370–373.
- Patzwahl DR, Treue S. 2009. Combining spatial and feature-based attention within the receptive field of MT neurons. *Vision Res.* 49:6.
- Shih JY, Atencio CA, Schreiner CE. 2011. Improved stimulus representation by short interspike intervals in primary auditory cortex. *J Neurophysiol.* 105:1908–1917.
- Treue S. 2001. Neural correlates of attention in primate visual cortex. *Trends Neurosci.* 24:295–300.
- 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. 1996. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature.* 382:539–541.
- 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.
- Wiederman SD, O'Carroll DC. 2013. Selective attention in an insect visual neuron. *Curr Biol.* 23:156–161.
- Womelsdorf T, Ardid S, Everling S, Valiente TA. 2014. Burst firing synchronizes prefrontal and anterior cingulate cortex during attentional control. *Curr Biol.* 24:2613–2621.

Supplementary figures

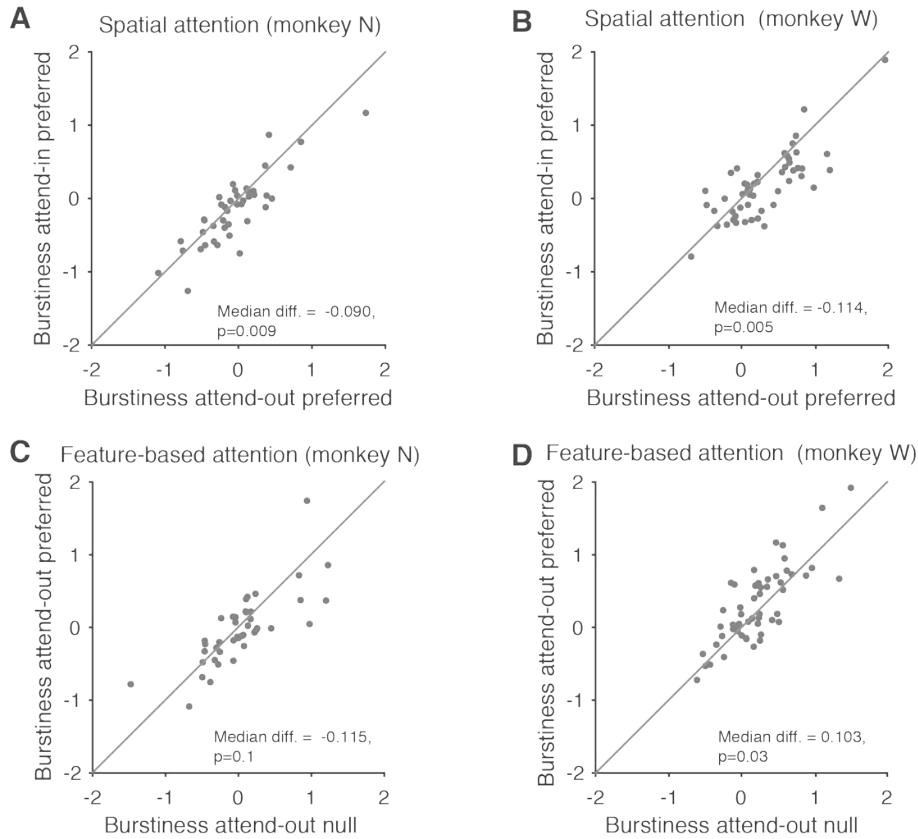


Fig. S1. Scatterplots showing the burstiness effects for the two monkeys: spatial attention reduces burstiness in both monkeys (A,B) while feature-based attention significantly increases burstiness in one monkey (C,D). Each data point represents one neuron and the p-value is the result of a signed-rank test. The median difference (median diff.) between the ordinate and abscissa ($y-x$), and the p-value from a signed rank test are indicated at the bottom-right of each panel.

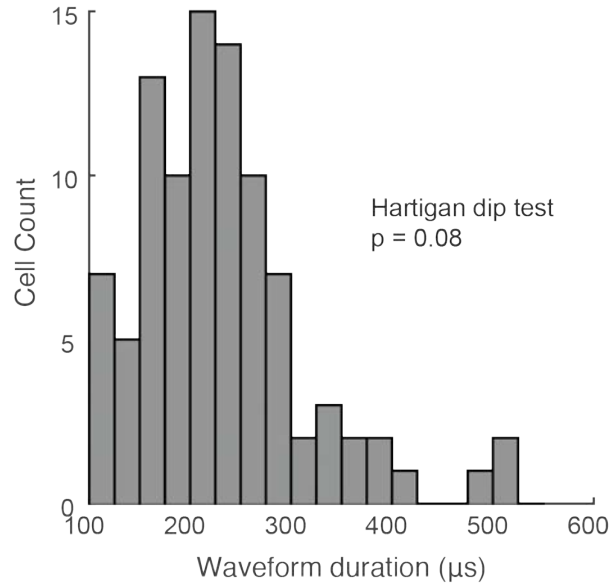


Fig. S2. No evidence for a bimodal distribution of waveform durations in our dataset. Histogram showing the distribution of waveform durations for all the recorded neurons. The p-value from a Hartigan dip test on the distribution is reported in the panel (same test for monkey W, $p = 0.4$ and for monkey N, $p \sim 1$).

List of cells included in analysis:

Cell	Channel		
dan-FMN2-nic-119-02-01_CV	2	dan-LFA1-wal-080-01-01_CV	1
dan-FMN2-nic-119-01-01_CV	1	dan-LFA1-wal-079-01-01_CV	1
dan-FMN2-nic-119-02-01_CV	10	dan-LFA1-nic-109-02-01_CV	1
dan-FMN2-wal-101-01-01_CV	1	dan-LFA1-wal-077-01-01_CV	1
dan-LFA1-nic-101-01-01_CV	1	dan-LFA1-wal-074-02-01_CV	1
dan-LFA1-nic-103-02-01_CV	2	dan-LFA1-wal-074-01-01_CV	1
dan-LFA1-nic-104-02-01_CV	1	dan-LFA1-wal-073-02-01_CV	3
dan-LFA1-nic-105-01-01_CV	1	dan-LFA2-nic-114-01-01_CV	1
dan-LFA1-nic-105-02-01_CV	1	dan-LFA2-nic-115-01-01_CV	3
dan-LFA1-nic-106-01-01_CV	1	dan-LFA2-nic-116-01-02_CV	1
dan-LFA1-nic-106-01-01_CV	1	dan-LFA2-nic-118-01-01_CV	10
dan-LFA1-nic-109-01-02_CV	2	dan-LFA2-nic-118-02-01_CV	1
dan-LFA1-nic-109-02-01_CV	1	dan-LFA2-nic-118-02-01_CV	9
dan-LFA1-nic-112-02-01_CV	1	dan-LFA2-wal-095-01-01_CV	1
dan-LFA1-nic-113-01-01_CV	1	dan-LFA2-wal-100-01-01_CV	9
dan-LFA1-nic-113-02-01_CV	3	dan-LFA2-wal-100-01-01_CV	1
dan-LFA1-wal-079-01-01_CV	1	dan-LFA2-wal-099-02-01_CV	1
dan-LFA1-wal-078-01-01_CV	2	dan-LFA2-wal-098-01-01_CV	1
dan-LFA1-wal-081-01-01_CV	1	dan-LFA2-wal-097-01-02_CV	2

dan-LFA2-wal-096-02-01_CV	9	dan-RFA1-wal-052-01-01_CV	1
dan-LFA2-wal-096-02-01_CV	1	dan-RFV2-nic-122-01-02_CV	2
dan-LFA2-wal-094-01-02_CV	1	dan-RFV2-nic-122-02-01_CV	9
dan-LFA2-wal-092-01-01_CV	2	dan-RFV2-nic-123-02-01_CV	3
dan-LFA2-wal-093-01-01_CV	1	dan-RFV2-wal-118-02-03_CV	1
dan-LFA2-wal-091-01-01_CV	1	dan-RFV2-wal-117-01-01_CV	1
dan-LFA2-wal-087-02-01_CV	3	dan-RFV2-wal-116-01-01_CV	2
dan-LFA2-wal-086-01-01_CV	2	dan-RFV2-wal-116-01-01_CV	2
dan-LFA2-wal-084-01-01_CV	2	dan-RFV2-wal-115-01-02_CV	2
dan-LFA2-wal-083-01-01_CV	1	dan-RFV2-wal-114-01-01_CV	3
dan-MFA1-nic-073-01-01_CV	1	dan-RFV2-wal-113-02-01_CV	1
dan-MFA1-nic-079-01-01_CV	1	dan-RFV2-wal-113-01-01_CV	1
dan-MFA1-nic-076-01-02_CV	1	dan-RFV2-wal-111-02-01_CV	1
dan-MFA1-nic-077-02-01_CV	1	dan-RFV2-wal-111-01-02_CV	1
dan-MFA1-wal-050-02-01_CV	3	dan-RFV2-wal-110-01-01_CV	1
dan-MFA1-wal-050-01-01_CV	3	dan-RFV2-wal-109-01-01_CV	1
dan-MFA1-wal-049-01-01_CV	1	dan-RFV2-wal-108-02-01_CV	3
dan-RFA1-nic-080-01-01_CV	2	dan-RFV2-wal-108-01-01_CV	1
dan-RFA1-nic-081-01-02_CV	1	dan-RFV2-wal-107-01-01_CV	2
dan-RFA1-nic-082-02-02_CV	3	dan-RFV2-wal-107-02-01_CV	2
dan-RFA1-nic-084-01-01_CV	3	dan-RFV2-wal-105-01-01_CV	1
dan-RFA1-nic-090-01-01_CV	1	dan-RFV2-wal-104-01-01_CV	1
dan-RFA1-nic-091-02-01_CV	1	dan-RFV2-wal-102-01-02_CV	1
dan-RFA1-nic-092-01-02_CV	1	dan-RFV3-nic-125-01-05_CV	1
dan-RFA1-nic-094-01-01_CV	2	dan-RFV3-nic-125-01-05_CV	3
dan-RFA1-nic-095-01-01_CV	1	dan-RFV3-nic-126-01-01_CV	1
dan-RFA1-nic-098-02-01_CV	1	dan-RFV3-nic-129-01-01_CV	3
dan-RFA1-nic-098-03-01_CV	1	dan-RFV3-nic-129-01-01_CV	11
dan-RFA1-nic-099-01-01_CV	1	dan-RFV3-wal-127-01-01_CV	3
dan-RFA1-nic-099-02-01_CV	1	dan-RFV3-wal-126-01-01_CV	10
dan-RFA1-wal-070-01-01_CV	1	dan-RFV3-wal-125-02-01_CV	2
dan-RFA1-wal-061-01-01_CV	1	dan-RFV3-wal-125-01-01_CV	3
dan-RFA1-wal-060-01-01_CV	1	dan-RFV3-wal-124-01-01_CV	1
dan-RFA1-wal-057-02-01_CV	1	dan-RFV3-wal-123-01-01_CV	3

Sustained spatial attention accounts for the direction bias of human microsaccades

Cheng Xue^{1,*}, Antonino Calapai^{1,2,*}, Julius Krumbiegel³, Stefan Treue^{1,2,3}

1 Cognitive Neuroscience Laboratory, German Primate Center, Goettingen, Germany

2 Leibniz-ScienceCampus Primate Cognition, Goettingen, Germany

3 Faculty of Biology and Psychology, Goettingen University, Goettingen, Germany,

* These authors contributed equally to this work.

Abstract

Microsaccades are involuntary small eye movements that happen while we maintain our gaze on a stationary point. Previous studies have shown that shortly after a symbolic spatial cue, indicating a behaviorally relevant location, microsaccades tend to be directed toward the cued region. This has led to the theory that microsaccades can be seen as an index for the covert orientation of spatial attention. However, this hypothesis faces two major issues. First, physiological effects of visual spatial attention are entangled with those of saccade planning. In this respect a systematic investigation is needed to assess to which extent saccade planning can influence microsaccade directions. Second, it is unclear whether the observed microsaccade direction effect is attention-specific or rather cue-specific. To address the first issue, we investigated the direction of microsaccades in human subjects when they attend to a behaviorally relevant location, while preparing a

response eye movement either toward or away from this location. We find that directions of microsaccades are in fact biased toward the attended location rather than towards the saccade target. To tackle the second issue, we verbally instructed the subjects about the location to attend, before the start of each block, so as to exclude potential visual cue-specific effects on microsaccades. Results indicate that despite the absence of visual cues during the experiment, sustained spatial attention alone reliably produces the microsaccade direction effect. Overall, our findings demonstrate that sustained spatial attention, without influences from saccade planning or the spatial cue per se, is sufficient to explain the direction bias observed in microsaccades.

Introduction

Microsaccades are involuntary, small ballistic eye movements that occur during gaze fixation¹. They have long been considered as noise in the eye movement system² until research within 15 years revealed some non-trivial feature about their frequency and direction. It has been reported in several human psychophysical studies, that around 300 ms after subjects are instructed by a symbolic spatial cue (e.g. an arrow-head at gaze location, a pre-assigned color or a sound source) to attend to a certain location, the directions of microsaccades were biased toward the location indicated by the cue, suggesting microsaccade's role as an index for covert spatial attention³⁻⁵. However, such an attention-specific interpretation of the post-cue microsaccade direction bias faces two challenges. First, while visual spatial attention is known to be closely entangled with saccade planning^{6,7}, such planning is known to interfere with the dynamics of

microsaccades⁵. Thus, to truly attribute the microsaccade direction effect to attention, it is necessary to remove any effect of saccade planning. Secondly, it is unclear whether the microsaccade direction effect is a reliable index of sustained attention, or, alternatively, merely a transient effect⁸. Specifically, previous studies that report the post-cue microsaccade direction bias have focused on a very specific time window, around 300 ms after the cue onset^{3,5,8,9}, and little to no evidence is available regarding whether this effect would last as long as spatial attention is maintained, or if it is only triggered by an immediately preceding spatial cue. It is important to note that for exogenous cues (i.e. a visual stimulus at the cued location) directions of microsaccades are directed away from the cued location^{5,8}, in contrast to the effect induced by an endogenous one.

To address these challenges, we recorded human eye movements during periods of fixation while the subjects performed a spatial attention guided *match to sample* task (Figure 1A). Our results demonstrate a consistent spatial attention effect on microsaccade directions, that is not directly triggered by a spatial cue, free from influences of saccade planning. These findings not only show the tight correlation between microsaccade direction and the subjects' internal attentional state, but also challenge the notion that spatial attention is functionally equivalent to a planning process of unexecuted movement (i.e. the premotor theory of attention⁶).

Methods

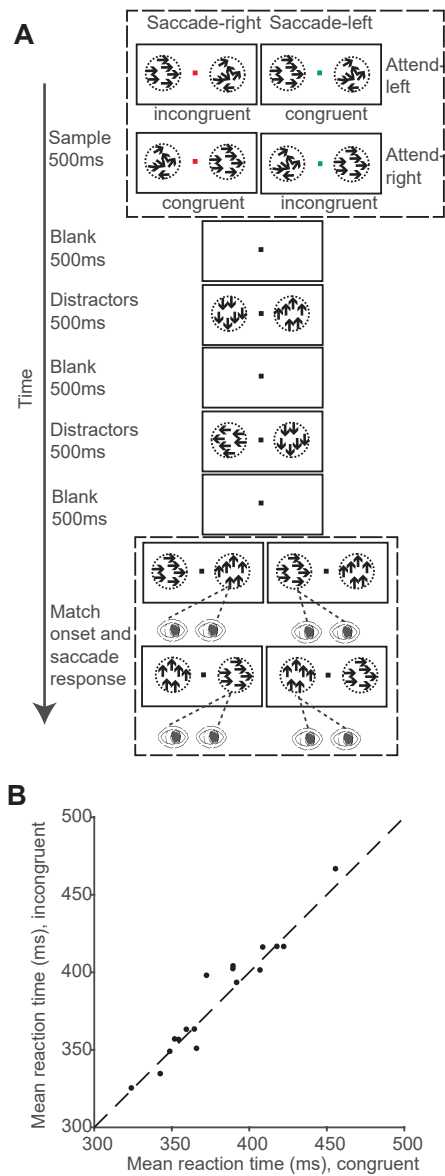


Figure 1. Match to sample task to dissociate attention and saccade planning. A) Task flow. Once the subject pressed a button and foveated the central fixation point. One fully coherent RDP and one non-coherent RDP were displayed. The coherent RDP is the sample stimulus. After a brief blank interval, a series of stimuli-pairs followed, and the subjects needed to respond when they found a match with the sample, and otherwise maintain fixation. The match can occur in any stimuli-pair at the same location as the sample, or in a small fraction of trials, does not appear at all. When the subjects found the match, they have to respond by making a saccade to one of the stimulus locations, which was instructed by the color of the fixation dot during the sample phase (red for rightward saccade, green for leftward saccade). B) Mean reaction times of incongruent hit trials (when the match appeared and the subjects correctly responded) plotted against that of congruent hit trials. Each dot represents one human subject. The dashed diagonal line indicates unity line.

Experiment setup

For both experiments, participants were seated at 57 cm distance from a 22" Samsung SyncMaster 2233RZ monitor, operating at a resolution of 1680 x 1050 pixels, with 120 Hz refresh rate. Eye Movements were acquired with an Eyelink 1000 (Version 4.56) while each subject's chin rested on a platform to maintain head position throughout the experimental sessions. The open-source software MWorks (Version 0.5) was used to run the tasks and to record the subjects' behavioral data.

Human subjects

This study recruited 35 naïve subjects (16 for experiment 1, 19 for experiment 2), whose gender, age, handedness, and vision profiles were listed in supplementary table 1. The study was approved by the Ethics Committee for experiments with humans of the Georg-Elias-Müller-Institute of Psychology, University of Göttingen, and followed the principles of the Declaration of Helsinki. Each subject received verbal and written information about the task, and gave written consent before the experiment started, and received monetary compensations after the experiment.

Experiment 1

In experiment 1 (Figure 1A), subjects depressed a button on a game pad (Logitech Inc., Precision) to start a trial. During the trial, subjects were required to maintain eye fixation at a central dot (size = 1 degree of visual angle – *dva* – in diameter, luminance = 5.65 cd/m², fixation window 2 *dva* in radius) until they decided to make a saccade to the required goals

as a response. Other fixation breaks would terminate the trial, which would be repeated later. Upon trial start, the fixation dot took on a color (either red or green) that informed the subjects about the way of response at the end of the trial (by making a rightward or a leftward saccade). During a sample phase, one random dot motion pattern (RDP, size = 8 dva; luminance = 30.09 cd/m², number of dots = 100, dot size = 0.25 dva, speed 5 dva/second) were displayed in each visual hemifield (RDPs centered 15 dva from the fixation point). One RDP had dots moving in random directions with zero coherence, and was irrelevant for the behavioral task. The other RDP (the sample) had coherently moving dots in one of four cardinal directions (up, right, down, left). The sample was followed by up to three alternating blank periods and displays of fully-coherent RDP-pairs. Subjects were required to detect a RDP with the same motion direction with the sample (a *match* stimulus), which might appear in any stimulus display period. In 10% of the trials, none of the three stimulus display periods contained a match, in which case the subjects just needed to maintain fixation till the end of the trial. The match, if it appeared, would always be at the same location as the sample. To report a match-detection, the subjects needed to make a saccade (either leftward or rightward) according to the color of fixation dot during the sample phase (green or red). Therefore, the response-saccade can be directed towards the same side as the match (a pro-saccade in a congruent trial), or to its opposite side (an anti-saccade in an incongruent trial). After match appearance, the subjects was required to respond within a time window individually determined for each subject through a staircase procedure prior the experiment started. The subjects performed the trials (with auditory feedback about the trial outcome at the end of each trial) as the response time

window adapted, until their performance stabilized at 80%; and the corresponding response time window was used throughout the following experiment. Each subject needed to correctly perform 480 trial to complete the experiment.

Experiment 2

A total of 19 subjects took part in Experiment 2. The task for the subjects was similar to experiment 1 except two major distinctions: (1) the trials were performed in blocks (80 correctly performed incongruent trials each block); within each block, all trials had a fixed location of attention (left or right), and a fixed goal for response-saccades (always on the other visual hemifield of the location of attention); (2) the sample phase contained only one fully-coherent sample stimulus located at the center. The location of attention (left or right) was instead given by a verbal instruction before each trial-block; while the goal for response-saccade in that block was inferred since all trials were incongruent trials. In other words, no stimulus during a trial block was spatially informative in any way. Stimuli used in experiment 2 were similar to that of experiment 1: fixation dot, 0.5 dva in diameter, luminance = 59.91 cd/m²; RDPs, size = 8 dva; luminance = 30.09 cd/m², number of dots = 100, dot size = 0.15 dva, speed 4 dva/second.

Microsaccade detection

We adopted the commonly used velocity threshold method described in Engbert & Kliegl, 2003 for microsaccade detection. We calculated the velocity for each eye at each millisecond based on the measured eye positions within a shifting time window of 8

milliseconds. The velocity threshold for each eye is then set at six times the standard deviation of all velocity magnitudes. All threshold crossing events are then compared between the two eyes, and only those with binocular threshold crossings are marked as microsaccades³. We detected 10426 microsaccades in experiment 1, and 6790 microsaccades in experiment 2. The algorithm-detected microsaccades were also visually inspected, and the start or endpoint of 643 microsaccades were manually corrected. This operation does not affect the directions of those microsaccades, which were defined as the direction of the peak velocity of the microsaccade.

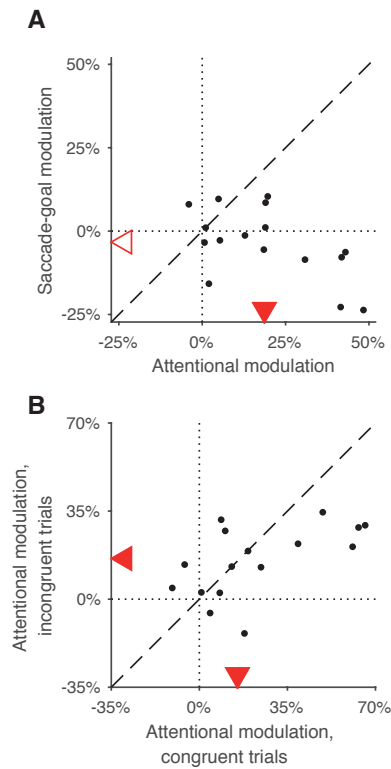


Figure 2. Overall microsaccade-directional modulation (see Material and Methods). A) The microsaccade-directional modulations by attended location (abscissae) plotted against microsaccade-directional modulations by saccade goal (ordinates). Each dot represents one subject. For both attended location and saccade goal, a positive modulation indicates a microsaccade-directional bias towards the respective location. Dotted vertical and horizontal lines indicate the zero line of abscissa and ordinates, respectively. Dashed line shows the unity line. Red arrows on horizontal and vertical axes indicate the median of abscissa and ordinates among all subjects, respectively. Filled symbols indicate the median is significantly different from zero; while open symbols, not. B) The microsaccade-directional modulation by attention for congruent-cue trials (abscissa) plotted against that for incongruent-cue trials (ordinates). Lines and symbols are similarly defined as in A).

This detection procedure clearly distinguished microsaccade from other smaller fixational eye movements and potential noise in the measurement (Figure S1, example microsaccade traces). The detected microsaccades showed a linear relationship (Pearson's correlation coefficient = 0.94, $p < 0.0001$) between amplitude and maximal speed (also known as the main sequence¹⁰, see Figure S2). We also observe that after a change in visual stimuli (e.g. the offset of stimuli), the rate of detected microsaccades temporarily drops, and rises to a peak at around 250-300ms after the stimulus change (Figure S3): a similar observation to what had been reported in many other studies³⁻⁵.

To investigate the microsaccade-directional profile while the subjects were expecting a potential upcoming match, most results were based on microsaccades that occurred during the blank periods (except in Figure 3, where direction profile were compared during stimuli with that during blank).

Results

Subjects were required to respond to the onset of a certain stimulus at one of two locations. We looked into the effect of the behaviorally relevant location on the distribution of microsaccade-directions, and whether such an effect is contingent on saccade planning or spatial cuing.

Simultaneous attentional deployment and saccade preparation

To disentangle the effects of spatial attention and saccade planning on microsaccade-direction, two independent spatial cues were given at the beginning of each

trial in experiment 1: the attention cue, indicating the location (left or right side of the screen) of the match if it appears, and the saccade cue, instructing the goal of the response-saccade (towards left or right). The locations of both cues were randomized for each trial, indicating either the same location (congruent trials) or opposite locations (incongruent trials). By dividing the trials either according to the location of spatial attention or the goal of response saccade, the influences of spatial attention and saccade planning on microsaccade direction can be separately evaluated.

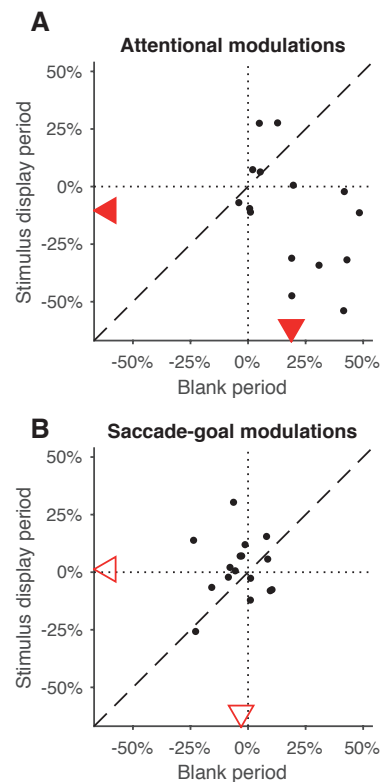


Figure 3. Microsaccade-directional modulations during blank periods versus stimulus display periods. A) shows the modulations by attended location, B) shows the modulations by saccade goal. In both A) and B), each dot represents one subject; its abscissa and ordinate represent the microsaccade-directional modulations during blank periods and during stimulus display periods, respectively. Dotted vertical and horizontal lines indicate the zero line of abscissa and ordinates. Dashed line shows the unity line. Red arrows on horizontal and vertical axes indicate the median of abscissa and ordinates among all subjects, respectively. Filled symbols indicate the median is significantly different from zero; while open symbols, not.

One critical objective of the experimental design is to encourage the subjects to plan a saccade to a given location already before the match appears (while also attending to an

independent location), rather than to plan a saccade only after match detection (when the attended location is no longer relevant). Given that the subjects are under time pressure to respond as quickly as possible (see methods), the latter strategy would likely lead to a longer reaction time in incongruent trials than in congruent trials. However, none of our 16 subjects showed significantly different reaction times between the two trial types (Bonferroni-Holm corrected rank sum test, $p > 0.05$ for all subjects) Figure 1B shows the subjects' mean reaction times for congruent trials (abscissa of the scattered dots), and for incongruent trials (ordinates of the scattered dots), respectively. There is no significant pair-wise difference across subjects ($p = 0.8$, Wilcoxon signed rank test), either.

Microsaccade-direction is biased toward the attended location, not the saccade goal. For each subject, we compare the distributions of microsaccade-directions during *attend left* trials against *attend right* trials. Taking the difference of leftward-microsaccade proportions of the two trials types provides a quantitative measure for the magnitude of attentional modulation of microsaccade directions: a positive attentional modulation indicates a bias of microsaccade direction towards the attended location, while a negative indicates a bias away from it. The abscissa of the scatter plot Figure 2A show the microsaccade-directional modulations by spatial attention for the 16 subjects tested in this experiment. The directions of microsaccades were significantly biased towards the attended location (, median 18.71%, $p = 0.0009$, Wilcoxon signed-rank test).

Similarly, by taking the difference of leftward-microsaccade proportions of trials in which

the response saccades are directed towards the left or the right hemifield, we determined the microsaccade-directional modulation by saccade-goal locations. The ordinates of the scatter plot Figure 2A show the microsaccade-directional modulations by saccade-goal. We did not observe any significant effect of the planned saccade goal on the direction of microsaccades (median -3%, $p=0.3$, Wilcoxon signed-rank test). A pair-wise signed-rank test also confirms that the attentional modulations are significantly larger than saccade-goal modulations, both toward the saccade-goal (comparing attentional modulations and saccade-goal modulations, $p=0.003$, Wilcoxon signed-rank test), and away from the saccade goal (comparing attentional modulations and the reversed saccade-goal modulations, $p=0.003$, Wilcoxon signed-rank test). There is also no significant correlation between attentional modulations and saccade goal modulations (Kendall's rank correlation = -0.283, $p=0.1$).

Attentional effect on microsaccade direction is consistent for congruent and incongruent cue trials

Although the saccade-goal is not a significant modulatory factor of microsaccade directions, it could still have a significant interaction with attention. We therefore looked at the attentional modulations in congruent-cue trials (attention cue and saccade cue at the same location, shown with abscissa of the scatter plot Figure 3B) and incongruent-cue trials (attention cue and saccade cue at opposite locations, shown with ordinates of the scatter plot Figure 3B), respectively. Similar attentional modulations were observed in both trial types (congruent-cue trials, median 15.4%, $p=0.003$; incongruent-cue trials,

median 16.0%, $p=0.003$; Wilcoxon signed-rank test). There is also no significant difference between the sizes of attentional modulations ($p=0.2$, Wilcoxon signed-rank test). We also did a two-way ANOVA on the left microsaccade proportions in all four combinations of attention and saccade-goal locations, which also confirmed the above conclusions: attended location is a significant factor ($p<0.0001$), while saccade-goal is not ($p=0.8$), neither is the interaction between attention and saccade goal ($p=0.4$).

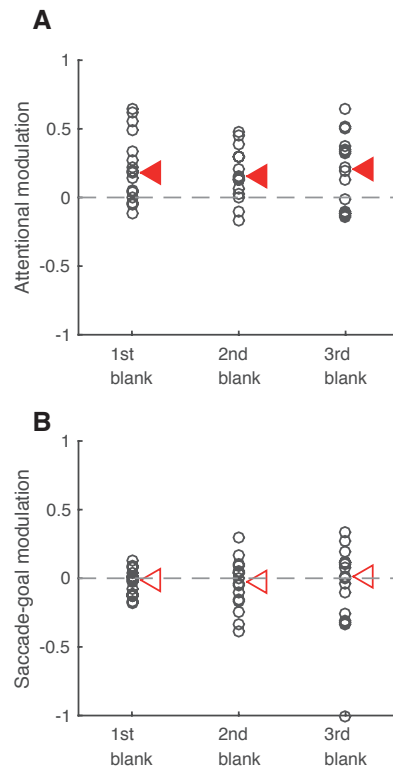


Figure 4. Microsaccade-directional modulations over the course of a trial. A) shows the modulations by attended location, B) shows the modulations by saccade goal. In both A) and B), directional modulations for microsaccades that occurred during the first, second, and third blank periods are shown separately for all subjects (denoted by circles). Dashed horizontal line indicates zero modulation. Red arrows indicate the median modulation during each blank period. Filled symbols indicate the median is significantly different from zero (Bonferroni-Holm corrected); while open symbols, not.

Attention modulates microsaccade-directional modulations during blank period and stimulus display period differently.

We have reported that attention biased microsaccade-direction towards the attended

location when the subjects were expecting the onset of a potential match, as shown by the abscissa of the scatter plots Figure 2A and Figure 3A.. Interestingly, however, when we look into microsaccade-directions during the display of RDP-pairs (i.e. Figure 1A distractor periods, during which the subjects correctly maintained fixation), as shown by the ordinates of the scatter plot Figure 3A, microsaccade-directions were biased away from the attended location (median -10.32%, $p=0.04$, Wilcoxon signed rank test). A pair-wise comparison between attentional modulations of microsaccade-direction during stimulus display periods and those during blank periods also showed significant difference ($p=0.01$, Wilcoxon signed rank test). Similarly, Figure 3B shows the saccade-goal modulations of microsaccade-directions during blank periods (abscissa) and during stimulus display period (ordinates). Saccade-goal does not have a significant effect on microsaccade-direction during stimulus display periods (median 1.3%, $p=0.6$, Wilcoxon signed rank test), not significantly different from its microsaccade-directional effects during the blank periods ($p=0.2$, Wilcoxon signed rank test).

Sustained attention, not spatial cue, modulates microsaccade-direction.

Previous studies have primarily reported a microsaccade direction effect around 300 ms after the spatial cue offset, when the microsaccade rate peaks. This makes it difficult to disentangle the role of sustained attention, and the role of the cue itself. In our design, the first blank period of each trial was preceded by the spatial cue (location of the sample), but the second and third blank periods were preceded with space-neutral distracting stimuli, which masked the direct visual influence from the attention cue. As shown in Figure 3A,

the attentional modulations for the three blank intervals were all positively shifted (first blank period, median 18.56%, $p=0.016$; second blank period, median 15.05%, $p=0.016$; third blank period, median 20.59%, $p=0.017$; all p values were calculated with Bonferroni-Holm corrected Wilcoxon signed rank test). This indicates that microsaccade-directions exhibit a significant bias toward the attended location, even if it is not immediately preceded by a spatial cue. Meanwhile, also consistent with the conclusions based on all blank period microsaccades, saccade-goal also does not significantly influence microsaccade-directions in any of the three periods alone (first blank period, median -1.93%, $p=0.7$; second blank period, median -2.23%, $p>0.9$; third blank period, median 0.77%, $p>0.9$; all p values were calculated with Bonferroni-Holm corrected Wilcoxon signed rank test).

To further isolate the microsaccade-directional modulation by sustained attention from potential confounds due to the visual stimulus used as a spatial cue, we conducted experiment 2, in which we tested the subjects with blocks of incongruent trials. All trials within each block had the same behavioral relevant location, and the same response saccade goal on the opposite side of the behavioral relevant location. Before each block started, we verbally gave the spatial cue, so that the trials within each block did not include a visual stimulus as spatial cue. We again calculated the attentional modulation of microsaccade directions by taking the difference between left microsaccade proportions in attend-left trial block and in attend-right trial block. The distribution of attentional modulations on microsaccade-directions is plotted in the histogram Figure 5. Despite the

absence of a spatial cue, we still found similar attention effects on microsaccade direction as in experiment 1 (median 8.9%, $p=0.003$), induced only by the subjects' prior knowledge of the location of the potential match. This strongly supports the hypothesis that sustained attention alone is enough to explain the bias in microsaccade direction.

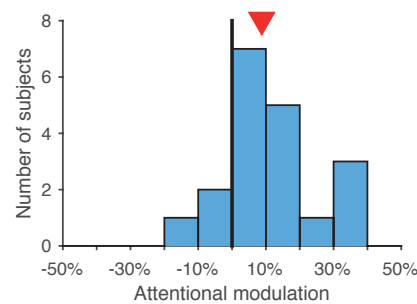


Figure 5. Microsaccade-directional modulations by endogenous attention, without preceding visual cue. Black vertical lines indicates no modulation. The red arrow over the histogram shows the median modulation. As in previously plots, filled symbol indicates the median is significantly different from zero.

Discussion

The exact interpretation over the nature of the microsaccade direction effect has been controversial. While many suggest that microsaccades can be considered as overt indicator of covert attention^{3-5,8,9,11-14}, others believe microsaccade-directions are dependent on other factors than spatial attention¹⁵⁻²⁰. Our study attempted to disentangle microsaccade-direction effect induced by attentional allocation from confounding factors, such as oculomotor preparation, and direct effect from visual cue. Our results showed that microsaccade-direction was biased toward attended location when the subject is expecting an upcoming target. Such an effect can be induced by sustained attention alone, and is not

contingent on an immediately-preceding spatial cue. However, we found no evidence for either a direct influence on microsaccades' direction from oculomotor planning, or an interaction between attention and oculomotor planning ¹⁴.

Spatial attention and oculomotor planning at the same time

Many studies have revealed the entanglement between visual spatial attention and oculomotor planning^{6,7,21}. Our way to separate the effects of the two is to encourage subjects to sustain spatial attention to one location, while at the same time also be ready to release a saccade towards an independent location. However, the pre-motor theory of attention, which posits that visual spatial attention is a result of oculomotor planning towards the attended location, would necessarily mean that subjects could only plan a saccade to a different location after their spatial attention is disengaged from a previous location, i.e. after the match is detected. Given that we ensure all subjects were under similar time pressure to respond as soon as possible (see staircase procedure in Methods), stimulus-response compatibility would predict longer reaction times for incongruent trials than for congruent trials ²², if oculomotor planning did not take place before the saccade go signal (the match onset). The fact that our subjects do not show such difference (either between reaction time distributions within each individual subject, or between mean reaction times across subjects) strongly suggests that oculomotor planning occurred before match onset, while the subjects were also expecting a potential match at an independent location.

Exogenous cue and inhibition of return

Rolfs and his colleagues reported that when using an exogenous cue to indicate a behavioral relevant location, microsaccades tend to be directed away from the cued locations⁵, an opposite effect than when an endogenous cue is used³, bringing questions upon whether the microsaccade-directional modulation is a visual cue-sensitive effect, rather than an attention effect. Meanwhile, it is hypothesized that this opposite effect of an exogenous cue could be attributed to the inhibition of return (IOR) caused by the onset of a salient peripheral visual stimulus ²³. Indeed, Galfano and colleagues showed that microsaccades were directed away from a peripheral stimulus that was not even behaviourally relevant ¹⁵. Besides Rolfs et al 2005, in other exogenously cued attention tasks, it seems such opposing effects on microsaccade-direction can also reach various equilibria, depending on stimulus-parameters: microsaccade-direction showed neither an IOR effect nor an attention effect, when the peripheral cue was merely a flash of small white dot within the context of a large symmetric stimulus-array¹⁸; or, microsaccade-direction can also show a net effect towards the cued location, when the peripheral cue was very close to the fixation point ⁴. In this study, we showed that when the peripheral attention cue was offset by a equiluminant stimulus at a symmetric location (experiment 1), or when attention was not instructed by a visual cue (experiment 2), microsaccade-direction was biased toward the attended location. In this context, our results provided further and stronger evidence that the microsaccade-directional modulation is a reliable attentional effect, not a cue-dependent effect. Meanwhile, the countereffect of IOR needs to be carefully controlled when studying

microsaccade-directional effect.

Different microsaccade-directional modulations during blank periods and stimulus display periods

Microsaccades are notoriously rare events³⁻⁵. By having alternating blank periods and stimulus display periods, our experiment is designed to induce many more microsaccades, to boost the statistical power of our analyses. On the other hand, one might wonder, with the relatively long trials in our experiment (up to 3.5 seconds) and microsaccades that are biased towards the attended location, how did the subjects ensure proper eye fixation? Our control analysis in Figure 3A showed that microsaccade-direction during stimulus-display periods was biased the opposite way to that during blank periods. This implies that microsaccades occurred during peripheral stimulus-display could have a distinct role: to correct eye position displacements^{11,24}. Since a match is not expected during the time a pair of distractor RDPs are still on display, It's conceivable that microsaccades during these periods reflect the oculomotor system's effort to countermand a fixation break after the sudden onset of peripheral RDP-pairs²⁵. However, to best address the role of microsaccades during stimulus display, future studies need to introduce stimulus-onset asynchrony to dissociate effects of stimulus-onset and potential confounding effects from expectation of future events.

The Premotor Theory of Attention: not the complete story?

There is much dispute whether saccade preparation and visual spatial attention share the

same neuronal mechanism, based on findings from various experiments, from human psychophysics to monkey neurophysiology^{6,26-28}. Our results shows: 1) incongruent trials do not take longer to complete than congruent trials; 2) when subjects are instructed to maintain spatial attention and also preparing for a saccade, microsaccadic directions are consistently biased towards the attended location, regardless of the saccade goal location. The former suggests that maintaining spatial attention and oculomotor planning can be executed at the same time toward different spatial locations. And the latter suggests that deploying spatial attention, but not preparing for saccades, has an effect on microsaccade directions. Given these findings, it is highly possible that there might be independent neuronal circuitries to implement visual spatial attention and oculomotor planning. Future monkey electrophysiology studies with a similar task might yield interesting insight into the neuronal basis of such mechanism.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center 889 “Cellular Mechanisms of Sensory Processing” and the Research Unit 1847 “Physiology of Distributed Computing Underlying Higher Brain Functions in Non-Human Primates”. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript. The authors thank Dr. Igor Kagan for helpful comments on this study and Kristin Dannhaeuser for help collect data in experiment 2.

Author contributions:

CX, AC and ST designed the experiment; CX and JK performed the experiment; CX analyzed the data; CX, AC and ST wrote the paper.

Conflict of Interest

The authors declare no competing financial interests.

References

1. Martinez-Conde, S., Macknik, S. L., Troncoso, X. G. & Hubel, D. H. Microsaccades: a neurophysiological analysis. *Trends in Neurosciences* **32**, 463–475 (2009).
2. Kowler, E. & Steinman, R. M. Small saccades serve no useful purpose: reply to a letter by R. W. Ditchburn. *Vision Research* **20**, 273–276 (1980).
3. Engbert, R. & Kliegl, R. Microsaccades uncover the orientation of covert attention. *Vision Research* **43**, 1035–1045 (2003).
4. Hafed, Z. M. & Clark, J. J. Microsaccades as an overt measure of covert attention shifts. *Vision Research* **42**, 2533–2545 (2002).
5. Rolfs, M., Engbert, R. & Kliegl, R. Crossmodal coupling of oculomotor control and spatial attention in vision and audition. *Exp Brain Res* **166**, 427–439 (2005).
6. Rizzolatti, G., Riggio, L., Dascola, I. & Umiltà, C. Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia* **25**, 31–40 (1987).
7. Smith, D. T. & Schenk, T. The Premotor theory of attention: Time to move on? *Neuropsychologia* **50**, 1104–1114 (2012).
8. Laubrock, J., Engbert, R. & Kliegl, R. Microsaccade dynamics during covert attention. *Vision Research* **45**, 721–730 (2005).
9. Laubrock, J., Kliegl, R., Rolfs, M. & Engbert, R. When do microsaccades follow spatial attention? *Attention, Perception, & Psychophysics* **72**, 683–694 (2010).
10. Bahill, A. T., Clark, M. R. & Stark, L. The main sequence, a tool for studying human eye movements. *Mathematical Biosciences* (1975).
11. Engbert, R. & Kliegl, R. Microsaccades keep the eyes' balance during fixation. *Psychological Science* **15**, 431–436 (2004).
12. Hermens, F. & Walker, R. What determines the direction of microsaccades? *Journal of Eye Movement Research* (2010).
13. Rolfs, M., Engbert, R. & Kliegl, R. Microsaccade orientation supports attentional enhancement opposite a peripheral cue: commentary on Tse, Sheinberg, and Logothetis (2003). *Psychological Science* **15**, 705–7– author reply 708–10 (2004).

14. Hermens, F., Zanker, J. M. & Walker, R. Microsaccades and preparatory set: a comparison between delayed and immediate, exogenous and endogenous pro- and anti-saccades. *Exp Brain Res* **201**, 489–498 (2009).
15. Galfano, G., Betta, E. & Turatto, M. Inhibition of return in microsaccades. *Exp Brain Res* **159**, 400–404 (2004).
16. Horowitz, T. S., Fine, E. M., Fencsik, D. E., Yurgenson, S. & Wolfe, J. M. Fixational eye movements are not an index of covert attention. *Psychological Science* **18**, 356–363 (2007).
17. Horowitz, T. S., Fencsik, D. E., Fine, E. M., Yurgenson, S. & Wolfe, J. M. Microsaccades and Attention: Does a Weak Correlation Make an Index?: Reply to Laubrock, Engbert, Rolfs, and Kliegl (2007). *Psychological Science* **18**, 367–368 (2007).
18. Tse, P. U., Sheinberg, D. L. & Logothetis, N. K. Fixational eye movements are not affected by abrupt onsets that capture attention. *Vision Research* (2002).
19. Tse, P. U., Sheinberg, D. S. & Logothetis, N. K. The Distribution of Microsaccade Directions Need Not Reveal the Location of Attention: Reply to Rolfs, Engbert, and Kliegl. *Psychological Science* **15**, 708–710 (2004).
20. Valsecchi, M., Betta, E. & Turatto, M. Visual oddballs induce prolonged microsaccadic inhibition. *Exp Brain Res* **177**, 196–208 (2007).
21. Moore, T. Microstimulation of the Frontal Eye Field and Its Effects on Covert Spatial Attention. *Journal of Neurophysiology* **91**, 152–162 (2003).
22. Munoz, D. P. & Everling, S. Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience* **5**, 218–228 (2004).
23. Hunt, A. R. & Kingstone, A. Inhibition of return: Dissociating attentional and oculomotor components. *Journal of Experimental Psychology: Human Perception and Performance* **29**, 1068–1074 (2003).
24. CORNSWEET, T. N. Determination of the stimuli for involuntary drifts and saccadic eye movements. *J Opt Soc Am* **46**, 987–993 (1956).
25. Tian, X., Yoshida, M. & Hamed, Z. M. A Microsaccadic Account of Attentional Capture and Inhibition of Return in Posner Cueing. *Front. Syst. Neurosci.* **10**, 428–23 (2016).
26. Corbetta, M. Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences* **95**, 831–838 (1998).
27. Craighero, L., Nascimben, M. & Fadiga, L. Eye Position Affects Orienting of Visuospatial Attention. *Current Biology* **14**, 331–333 (2004).
28. Moore, T. & Armstrong, K. M. Selective gating of visual signals by microstimulation of frontal cortex. *Nature* **421**, 370–373 (2003).

Supplementary material

Subject list, experiment 1. Note that handedness has been assessed verbally.

Subject	Age	Gender	Handedness	Vision
ANH	22	f	right	normal
CAM	20	f	right	normal
INB	21	f	right	contact lenses
INN	22	f	right	normal
JOD	23	m	right	normal
LEI	22	f	right	normal
MAL	26	m	right	normal
MIS	23	f	right	normal
AGN	23	f	right	glasses
SVE	23	f	right	normal
ANM	24	m	right	glasses
GAE	22	f	right	normal
ALN	20	f	right	normal
THR	23	f	right	normal
LUK	20	m	right	normal
MAG	27	f	right	normal

Subject list, experiment 2. Note that handedness has been assessed verbally.

Subject	Age	Gender	Handedness	Vision
FHA	23	m	right	normal
JKI	20	m	right	normal
MAS	25	m	left	contact lenses
JAK	24	f	right	contact lenses
PAK	24	f	right	contact lenses
LUB	25	f	right	normal
REP	23	m	right	glasses
KES	25	f	right	normal
SIL	33	f	right	normal
TEF	25	f	right	normal
RIW	23	f	right	normal
JUG	23	f	left	normal
DIT	22	f	left	contact lenses
CAR	21	f	right	contact lenses
SGO	21	f	right	normal
ANV	23	f	right	glasses
JMA	31	m	right	glasses

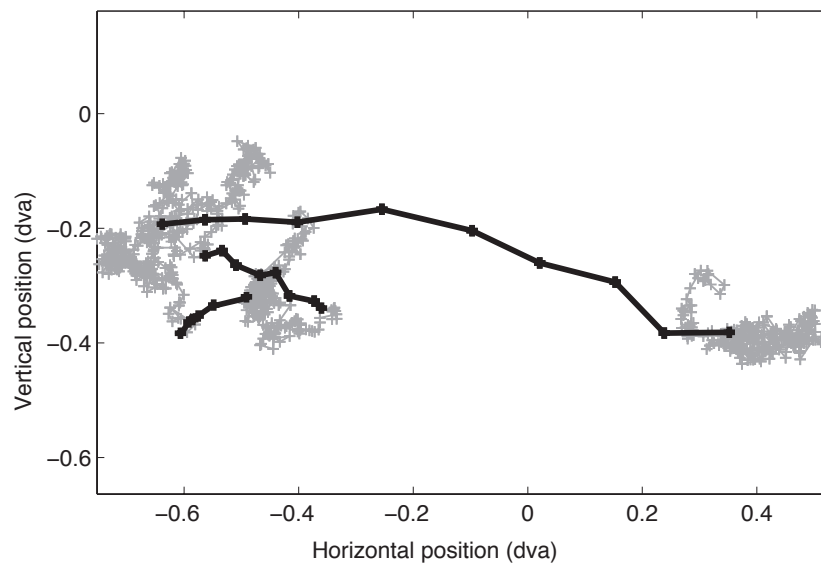


Figure S1. Example binocular eye trace. Black traces indicate detected microsaccades. Detected microsaccades clearly separates itself from other types of eye movement and system noise (grey traces.)

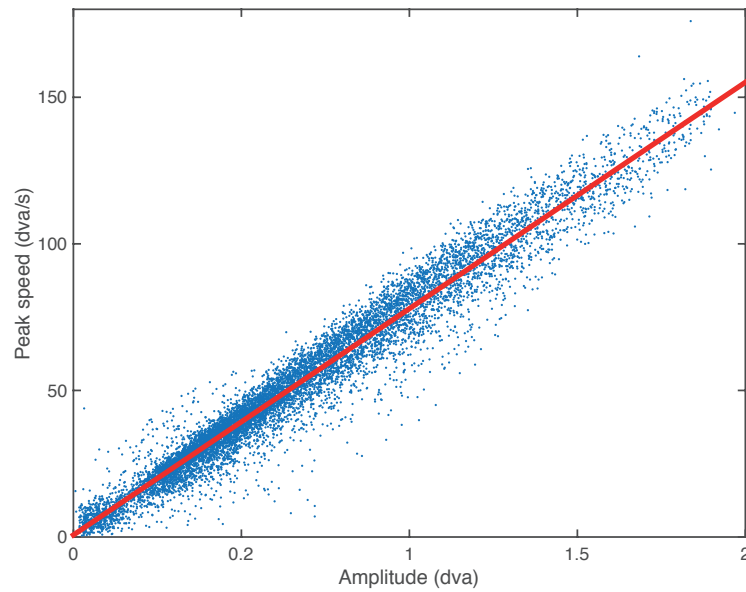


Figure S2. The main sequence. Abscissa and ordinate of the dots indicate the amplitude and peak speed of all the microsaccades across all subjects. black line indicate the linear fit of the data points.

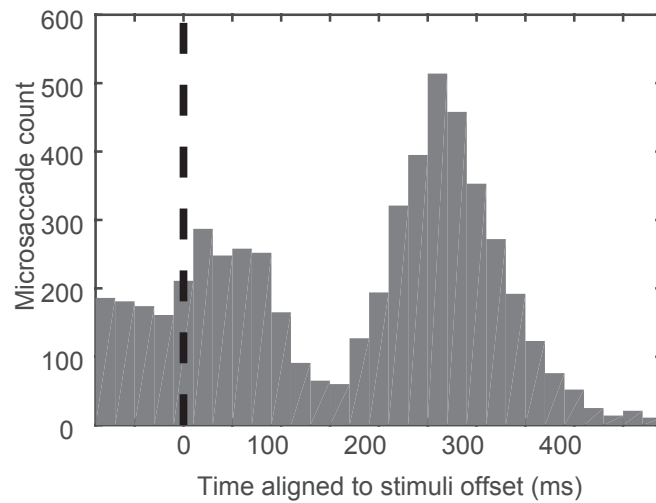


Figure S3. The distribution of microsaccades with respect to the stimulus onset, pooled across all subjects.



ORIGINAL ARTICLE

Differential Contribution of Low- and High-level Image Content to Eye Movements in Monkeys and Humans

Niklas Wilming^{1,2,3,4,9}, Tim C. Kietzmann^{1,5}, Megan Jutras^{2,3,9}, Cheng Xue⁶, Stefan Treue^{6,7,8}, Elizabeth A. Buffalo^{2,3,9} and Peter König^{1,4}

¹Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany, ²Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, USA, ³Yerkes National Primate Research Center, Atlanta, GA 30329, USA, ⁴Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, ⁵Medical Research Council, Cognition and Brain Sciences Unit, Cambridge CB2 7EF, UK, ⁶Cognitive Neuroscience Laboratory, German Primate Center - Leibniz-Institute for Primate Research, Goettingen, Germany, ⁷Faculty of Biology and Psychology, Goettingen University, Goettingen, Germany, ⁸Leibniz-ScienceCampus Primate Cognition, Goettingen, Germany and ⁹Washington National Primate Research Center, Seattle, WA 09195, USA

Address correspondence to Niklas Wilming. Email: nwilming@uke.de

Abstract

Oculomotor selection exerts a fundamental impact on our experience of the environment. To better understand the underlying principles, researchers typically rely on behavioral data from humans, and electrophysiological recordings in macaque monkeys. This approach rests on the assumption that the same selection processes are at play in both species. To test this assumption, we compared the viewing behavior of 106 humans and 11 macaques in an unconstrained free-viewing task. Our data-driven clustering analyses revealed distinct human and macaque clusters, indicating species-specific selection strategies. Yet, cross-species predictions were found to be above chance, indicating some level of shared behavior. Analyses relying on computational models of visual saliency indicate that such cross-species commonalities in free viewing are largely due to similar low-level selection mechanisms, with only a small contribution by shared higher level selection mechanisms and with consistent viewing behavior of monkeys being a subset of the consistent viewing behavior of humans.

Key words: human macaque comparison, low-level salience, oculomotor control, overt visual attention

Introduction

Eye movements are an essential aspect of our everyday behavior, because the direction of gaze determines what parts of our visual environment are processed with high-accuracy foveal vision. The importance of eye movements is reflected in their ubiquity (saccades occur at a rate of ca. 3–5 Hz) and in viewing strategies that are specifically tailored toward behavior (Land and Hayhoe 2001; Land and Tatler 2001; Sullivan et al. 2012;

Johnson et al. 2014). Understanding the underlying cortical saccade target selection process is therefore fundamental for understanding vision and human cognition at a larger scale (Petersen and Posner 2012).

The processes underlying such overt visual selection have traditionally been approached by behavioral measurements, mostly performed on humans, and by electrophysiology, performed in macaque monkeys, which are the most prominent

model system for studying attentional selection (Bisley 2011). This approach rests on the fundamental assumption that common neural mechanisms are at play in both species. Only then is the link of (monkey) neuronal mechanisms to (human) behavior valid. To verify this assumption, it is crucial to investigate whether the overall behavioral phenomenon to be understood, overt visual attention, is comparable in human and macaque. Here, we addressed this issue by comparing patterns of eye movements recorded from 11 monkeys and 106 human observers, while they were performing a task that comes natural to both species: free viewing. Free viewing has the advantage that it does not require explicit instructions or training. Furthermore, monkeys do not need to be externally rewarded because they are intrinsically motivated to freely explore visual scenes. Thus, free viewing can be completed without instructions, training, or explicit reward and it therefore remains undefined which parts of the stimuli should be attended. While tasks that require training can result in comparable behavior, they potentially mask the natural *modus operandi* of overt visual selection. Consequently, if free-viewing behavior is similar across humans and monkeys, it is because both species have intrinsically chosen a selection strategy that emphasizes the same locations, not because the task dictates which locations promise success. Free viewing therefore provides an unbiased view of the natural selection processes of overt attention in macaques and humans.

To compare viewing behavior across species, we followed a 2-staged approach. We first compared cross-species similarity in fixation locations. Using data-driven agglomerative clustering, we found that the 2 species form distinct clusters of viewing behavior, indicating species-specific selection strategies. Despite these differences, cross-species predictions were clearly above chance, indicating shared behavior. Following these observations, we tested in how far these differences and similarities in viewing behavior can be understood in terms of different explanatory dimensions, commonly assumed to jointly contribute to the guidance of eye movements. Distinctions are typically made between stimulus-dependent, context-dependent, and geometrical factors. Stimulus-dependent influences are, for example, the saliency conveyed by low-level image features (Itti and Koch 2001; Parkhurst et al. 2002), objects (Einhäuser et al. 2008; Nuthmann and Henderson 2010), and stimulus interpretation (Kietzmann et al. 2011). Context-dependent aspects include the task (Castelhano et al. 2009; Betz et al. 2010) and scene context (Torralba et al. 2006; Kietzmann and König 2015). Geometrical aspects include oculomotor biases, like the center bias of fixations (Tatler and Vincent 2009) and saccadic momentum (Smith and Henderson 2011; Wilming et al. 2013). All of these aspects interact in the selection process and consistently make strong contributions to the guidance of eye movements (Kollmorgen et al. 2010). These well-established dimensions therefore provide a good starting point to understand the observed similarities and dissimilarities across species. However, while low-level stimulus features are a well-controlled and well-studied explanatory dimension, higher level factors are less clearly defined in the context of free viewing on natural scenes, which comprises the current data set. We therefore initiate our investigation by comparing the relative contribution of low-level stimulus features and subsequently test any other residual, presumably higher level, factors across both species.

To estimate the relative contribution of these different factors, we first estimated the consistency of viewing behavior within humans and monkeys. The consistency within a species measures the similarity of viewing behavior across many observers and thereby forms an upper bound for the similarity of fixation selection strategies (Wilming et al. 2011). The reliability of such

consistency estimates depends on the number of observers (Wilming et al. 2011). In particular, small groups tend to underestimate the consistency within a group of observers and consistency estimates approach an asymptotic level as the group size increases. In this study, we compare 11 monkey and 106 human observers and, to our knowledge, our data set is the first to reach this asymptotic level. This analysis revealed an overall reduced consistency in macaques compared with humans. We then decomposed the respective upper bound into a stimulus-driven part and residual viewing behavior that must be driven by the remaining explanatory dimensions. These analyses revealed that the predictive power of low-level features is comparable across species. This implies that low-level features can explain large parts of the consistent macaque viewing behavior, but provide comparably limited predictive power in humans. However, the absolute impact of different low-level feature dimensions exhibits large similarities across species, suggesting that similar low-level selection mechanisms are at play in both macaques and humans. Following this observation, we tested whether commonalities across species can be observed beyond these, presumably shared, low-level mechanisms. We found that a joint model, combining low-level saliency and cross-species predictions, that is, data from humans to predict macaques and data from macaques to predict humans, only yields marginally better prediction accuracy than the low-level model alone. Thus, while our data suggest that human and macaque share common low-level selection mechanisms, other, potentially higher level effects only generalize to a small degree across species.

Materials and Methods

Participants

Eye movements were recorded from 11 rhesus monkeys (*Macaca mulatta*, 8 male). Recordings were carried out across 3 different locations. Data from 4 monkeys were recorded at the Yerkes National Primate Research Center (YNPRC) in Atlanta, USA, in accordance with National Institute of Health guidelines and protocols were approved by the Emory University Institutional Animal Care and Use Committee. Data from 3 additional monkeys were recorded at the Washington National Primate Research Center (WNPRC) in Seattle, USA, in accordance with National Institute of Health guidelines and protocols were approved by the Washington University Institutional Animal Care and Use Committee. Data from 4 additional monkeys were recorded at the German Primate Center (DPZ) in Goettingen, Germany, in accordance with European Directive 2010/63/EU, corresponding German animal welfare law and institutional guidelines. The animals were group-housed with other macaque monkeys. The facility provides the animals with an enriched environment (including a multitude of toys and wooden structures, natural as well as artificial light, exceeding the size requirements of the European regulations, including access to outdoor space; Calapai et al. 2016). All procedures were approved by the appropriate regional government office (Niedersächsisches Landesamt fuer Verbraucherschutz und Lebensmittelsicherheit, LAVES). Eye-movement recordings from humans came from 2 previous studies that used the same stimuli and comparable tasks. We analyzed data from 106 observers, 58 from Açik et al. (2010) and 48 from Onat et al. (2014). Açik et al. recruited participants from different age ranges (18 children with mean age 7.6 years, 6 female; 23 university students with mean age 22.1, 11 female; 17 older adults with mean age 80.6, 10 female). Onat et al. recruited 48 students (mean age 23.1 years, 25 male). The majority of participants were therefore recruited from the general

student population at the University of Osnabrück. All participants (main and control experiments) gave written informed consent and all experimental procedures for eye-movement recordings from humans were in compliance with guidelines described in the Declaration of Helsinki and approved by the ethics committee of the University of Osnabrück.

Stimuli

Stimuli consisted of 192 images from 3 different categories (64 images in each category). “Natural” scenes were taken from the “McGill Calibrated Color Image Database” and depict mainly bushes, flowers, and similar outdoor scenes. “Urban” scenes depicted urban and manmade scenes taken around Zürich, Switzerland. “Fractal” images were taken from Elena’s Fractal Gallery, Maria’s Fractal Explorer Gallery, and Chaotic N-Space Network available online, and depicted computer-generated fractals. Figure 1A shows example stimuli from all categories. Please see Açıık et al. (2010) for more details.

Apparatus

Recordings at the Yerkes National Primate Center and the Washington national primate center were carried out with an ISCAN infrared eye-tracking system while each monkey sat in a dimly illuminated room. Monkeys were head fixed during recordings. Stimuli were presented on a CRT Monitor with a resolution of 800×600 pixels and a refresh rate of 120 Hz. The viewing distance was 60 cm. Recordings at the German Primate Center were carried out in similar conditions but an SR-Research EyeLink 1000 was used for recording of eye movements. The viewing distance was 57 cm and stimuli were presented on a TFT screen (60 Hz, 1920×1080 pixels). The size of the images in degrees of visual angle was matched between all 3 setups ($33.3^\circ \times 25^\circ$).

Human eye movements were recorded with an EyeLink 1000 system (Açıık et al. 2010) or an EyeLink II system (Onat et al. 2014). Human eye-movement recordings were carried out at the University of Osnabrück, Germany. Onat et al. presented stimuli on a CRT Monitor with a resolution of 1280×960 pixels and a refresh rate of 85 Hz. The viewing distance was 80 cm.

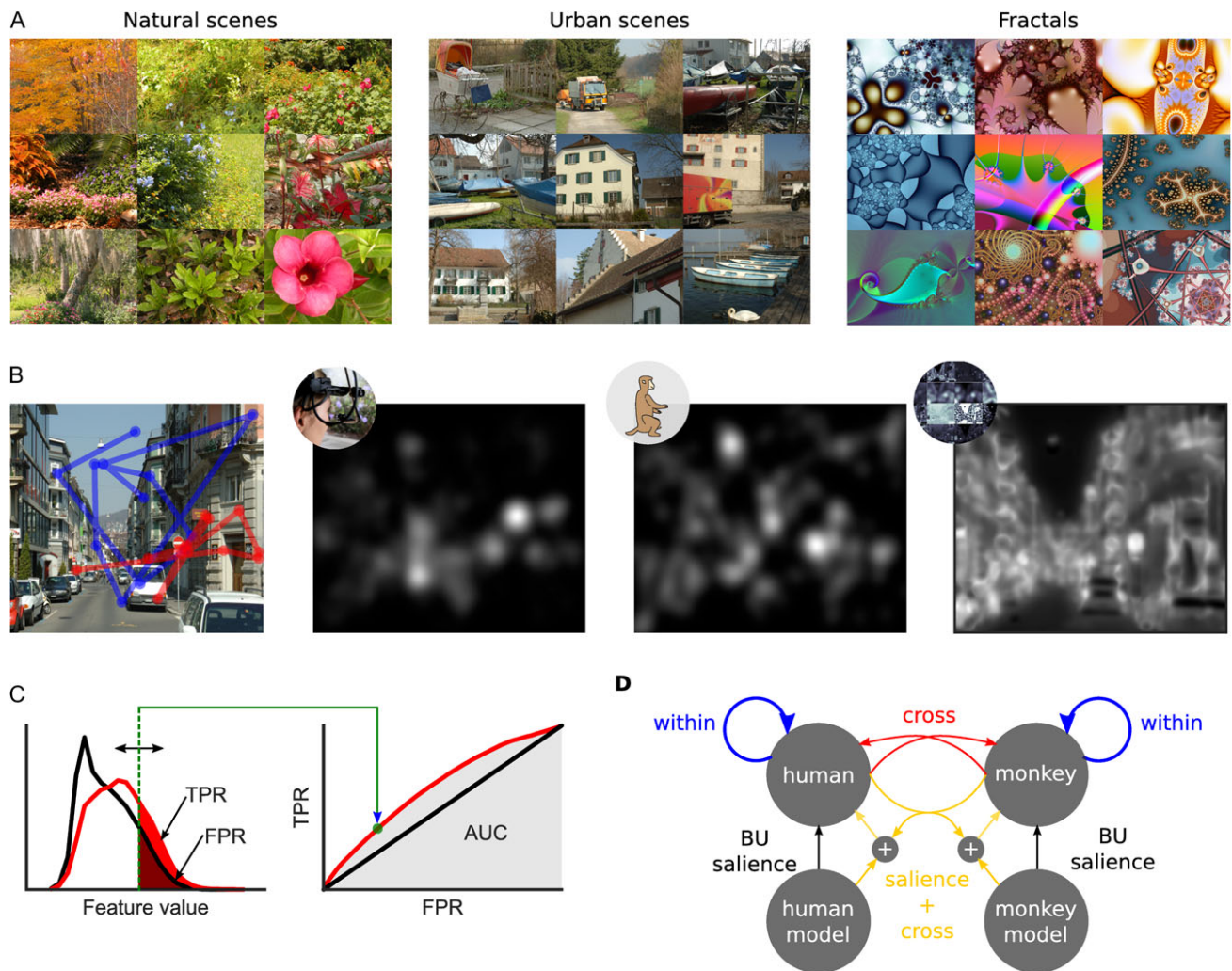


Figure 1. Study overview. (A) Nine example images from the categories natural scenes, urban scenes, and fractal scenes. (B) One example stimulus with one monkey (blue) and one human (red) eye-movement trace. The next 3 plots show the density of human fixations on the example image, the density of monkey fixations, and a predicted saliency map for the example stimulus. (C) The computation of AUC values. Left: Feature values at fixated locations (red) and non-fixated control locations (black) are classified as fixated or not fixated by a simple threshold (green dotted line). Moving the threshold and plotting the false alarm rate (FPR) against the hit rate (TPR) generates a receiver operating characteristic (ROC) curve which is shown on the right. The area under this curve (AUC) is a measure of classification quality. (D) Different predictors and comparisons in this study.

Açık et al. used a 60-Hz TFT screen with the same resolution and a viewing distance of 65 cm. The size of the images in degrees of visual angle was $35^\circ \times 26^\circ$ (Açık et al. 2010) and $30^\circ \times 22^\circ$ (Onat et al. 2014) for the human recordings.

Procedure and Task

Monkeys performed a free-viewing task and were not explicitly rewarded for image viewing. Images were shown until a total looking time inside the image of 10 s had accumulated. Monkeys at the Washington and Yerkes National Primate Center carried out a color change task between free-viewing trials. In this task, the monkey was required to hold a touch bar and maintain fixation on a small rectangle (0.3°) that appeared at various locations on the screen. The rectangle changed color from gray to an equiluminant yellow at a randomly chosen time between 500 and 1100 ms. Upon release of the touch bar within 500 ms after the color change, a drop of blended chow was delivered as reward (Jutras et al. 2009; Jutras and Buffalo 2010). Monkeys at the German Primate Center carried out a fixation control task that required them to saccade to a point on the screen and were rewarded for maintaining fixation for 1.25 s. Data from the control trials were not included in subsequent analyses.

Macaque recordings were carried out on 3 consecutive days. This kept sessions short enough for monkeys to attend to all images without losing interest. On the first 2 days, 66 randomly sampled images were shown twice and on the last day, 60 images were shown twice. The order of presentation was the same for all monkeys. Due to a technical error, the data from 1 day from one monkey was discarded. To increase the amount of available data, and to potentially compare effects of memory later on, 2 monkeys repeated the experiment after 4 weeks.

Human observers were instructed to “freely view” the same images for 6 s (Onat et al. 2014). In contrast, Açık et al. (2010) showed images for 5 s and instructed participants to “study the image carefully”. After each image, participants were then shown a 3.2° image patch and had to judge if it was taken from the image presented just before. We consider the 2 tasks to be comparable since the patch recognition task does not require special viewing strategies. In particular, patch locations were drawn uniformly from the entire image and patches are large and easily identifiable such that freely inspecting the image allows successful completion of the task. This was also reflected by the high task performance of the participants (85% across all age ranges). Both studies therefore used similar instructions, and the data were pooled accordingly. All analyses were performed on the first 5 s of image viewing.

Data Pre-processing

Saccade detection for humans was based on 3 measures: eye movement of at least 0.10° , with a velocity of at least $30^\circ/\text{s}$, and an acceleration of at least $8000^\circ/\text{s}^2$. After saccade onset, minimal saccade velocity was $25^\circ/\text{s}$. Saccade detection for monkeys was carried out similarly but we additionally required that each saccade lasted at least 21 ms and traveled at least 0.35° of visual angle. This was necessary to compensate for the lower sampling rate of the ISCANN system (240 Hz vs. 500 Hz and 1 kHz). Samples in between 2 saccades were labeled as fixations.

Monkey eye-tracking data recorded with the additional color change task were calibrated in 2 steps. Before each recording session, monkeys carried out a block of color change trials.

Since the color change was subtle, monkeys had to fixate the rectangle in order to detect the color change. We manually adjusted the gain and offset of the eye tracker until fixations were on the color change rectangle. To improve the manual calibration after the recording, we used the color change trials between picture presentations. We fitted a 2D affine transformation (least-squares fit) between average eye position after onset of the color change rectangle and the position of the rectangle in visual space. This took care of translations and skew in the monkey eye-tracking data. Monkeys from the German Primate Center were calibrated using a 12-point calibration grid before the task. Human eye tracking was calibrated with a 12-point grid before the experiment started.

Since the stimulus presentation time was different between experiments (5, 6, and 10 s), we only used the first 5 s of image viewing for subsequent analysis. We rescaled all eye-tracking data to the stimulus size used for monkeys in Atlanta and Seattle (800×600 pixel).

Performance Measure: Computation of AUC Values

This study investigated how well different factors predicted fixation locations of humans and monkeys. Specifically, we were interested in the predictive power of bottom-up-saliency, within-species consistency and cross-species consistency and fixation densities of individual observers. These factors were quantified by “predictors” (described in detail below) that assign a score to every location in an image, which scales with the predicted likelihood of fixating this location. To assess the quality of each predictor, we evaluated if fixated locations (“actuals”) received higher predictor scores than non-fixated control locations (“controls”). We computed the area under the receiver operating characteristic (ROC) curve (AUC), separating feature values at actual and control locations, as our performance measure.

The AUC is computed by classifying actual fixation locations and control locations as fixated or non-fixated based on the respective score at actual and control locations based on a simple threshold. Varying this threshold generated ROC curves for each predictor and the AUC is computed as the area under this ROC curve. The area sums to 1.0, if the classification is perfect, that is, the distributions of score values at actual and control locations are perfectly separated. A value of 0.5 indicates a classification at a chance level. Perfect misclassification results in an area under the curve of zero. To account for the center bias of fixations, control locations were drawn from the spatial bias of each observer (Tatler et al. 2005; Tatler 2007; Tatler and Vincent 2009). That is, control data were taken from the same subject on all other stimuli of the same category. Each predictor was evaluated for every observer and averaged over all stimuli within a category. Finally, we here aim at understanding the factors contributing to the consistent viewing behavior in each species. We will therefore express the predictive power of individual predictors relative to the within-species consistency, which serves as an upper bound. Since an AUC of 0.5 implies chance level performance, we subtract 0.5 from both AUC values before computing the ratio.

The within- and cross-species predictions consist of fixation densities that are generated by smoothing all fixations that form a predictor (e.g. all fixations on an image of one species for the cross-species predictor) with a Gaussian filter of FWHM = 2° and subsequently normalizing the 2D map to unit volume.

Consistency Between Individual Observers and Hierarchical Clustering

To investigate the similarity of viewing behavior of all pairs of observers, we computed AUC that indicate how well fixations from observer A on one stimulus predict the fixations from another observer B on the same stimulus. For each stimulus, we computed a fixation density map from fixations of observer A and computed how well the density values separate actual and control locations from observer B. Averaging across stimuli within a category yielded 3 similarity matrices with 117×117 AUC values each. These matrices are not symmetric since observer A and B have different control locations that affect the AUC value. We symmetrized the matrices by computing the average AUC of A predicting B and B predicting A, as required for the subsequent hierarchical clustering.

In a second step, we applied hierarchical agglomerative clustering with Ward's minimum variance criterion to each of the 3 category specific similarity matrices (Ward 1963). Agglomerative hierarchical clustering starts with individual observers and repeatedly merges them into clusters. In each iteration, clusters from the previous iteration (starting with individual observers) are merged such that the total within cluster variance increases as little as possible.

For technical reasons, we could not include all human observers in the clustering. AÇik et al. (2010) balanced stimulus presentation across observers, such that pairs of observers saw the entire data set. This implies that these pairs cannot predict each other since they did not see the same images. Agglomerative hierarchical clustering cannot easily deal with such missing values unless they are imputed, and the imputation strategy used can affect the clustering. To circumvent this ambiguity, we decided to use only a subset of the 117×117 similarity matrices, excluding observers that had missing values. Notably, however, the main results reported here hold true even if the additional observers are included, independent of the imputation strategy.

Cross- and Within-Species Consistency

How well fixation locations from one species predicted locations from the other was quantified with a cross-species predictor. Computing a fixation density with all fixations from one species on the image in question generated the cross-species prediction for a specific image. This yielded a score for each location in a visual stimulus, which was subsequently used to compute the cross-species AUC. The within-species prediction was similar, but used fixations from the own species (without the subject currently being evaluated). This within-species consistency forms an effective upper bound for the predictability of viewing behavior (Wilming et al. 2011).

The within-species consistency AUC is a biased estimator that tends to produce smaller values with fewer observers. To allow meaningful comparisons between humans and monkeys, we subsampled our human data set to fewer observers (see Statistical Comparisons).

Feature-Fixation Correlations

To compare the influence of image features on viewing behavior, we computed how well different features predicted fixation locations. In total, we computed 16 different features on 3 spatial scales. First, we computed luminance, blue-yellow, red-green, and saturation channels of all stimuli. A first group of

features represents a smoothed (Gaussian filter, $\sigma \in \{8 \text{ pixel}, 16 \text{ pixel}, 32 \text{ pixel}\}$) version of these simple features. A second group of features is computed by computing contrast in a Gaussian circular aperture ($\sigma \in \{8 \text{ pixel}, 16 \text{ pixel}, 32 \text{ pixel}\}$) on the luminance, blue-yellow, red-green, and saturation channels. Contrast was computed according to the following formula:

$$C = (G(X^2, \sigma) - G(X, \sigma)^2)^{1/2},$$

where G convolves the input X with a Gaussian kernel with standard deviation σ . The third group represents the second-level contrast maps (contrast of the contrast maps, $\sigma \in \{38 \text{ pixel}, 76 \text{ pixel}, 151 \text{ pixel}\}$), which describe texture contrast. A fourth group consists of an edge detection filter (Sobel filter) and the intrinsic dimensionality 0–2 of local image patches (cf. Onat et al. 2014). Intrinsic dimensionality describes how well a local patch is described by an edge, corner, or surface. Each feature was computed on 3 different spatial scales. This was achieved by down sampling of the original stimulus with a Gaussian pyramid up to 2 times. Each step of the pyramid halves the length and the width of the stimulus yielding the high (no down sampling), medium (once), and low (twice) spatial scale. We then computed the AUC for predicting fixation locations from each feature. This yielded a vector of 48 AUC values for each observer and category.

We then compared AUC values across monkey and human observers to investigate to what extent different features indicate similar predictive power, and therefore suggest similar selection mechanisms. AUC values < 0.5 indicate that a feature is anti-predictive of fixation locations, that is, would be predictive if feature maps were multiplied with -1 . This has an important implication for comparing features between species. The AUC value of features that are more predictive, for example, in monkeys, will be closer to one compared with the human feature AUC values. Those features that are more anti-predictive in monkeys will be closer to zero. If features are more effective in one species, independent of whether they are predictive or anti-predictive, feature AUCs will fall on a line that pivots around 0.5. In a regression that predicts monkey feature AUC values based on human feature AUC values such an increase in effectiveness would be demonstrated with a slope that is different from one. Regression coefficients larger than one indicate that features effective in humans are more effective in monkeys; coefficients smaller than one indicate that humans are more driven by bottom-up features.

We therefore repeatedly ($N = 1000$) regressed the pattern of average monkey feature-fixation AUCs onto an average of feature-fixation AUC vectors from 11 randomly sampled human observers. Subsampling AUC values from human observers allowed us to partially estimate the variance introduced by only having 11 monkey observers. Finally, we thus tested whether regression coefficients were different from one with a 2-sided t-test and computed the average variance explained by the regression models.

Salience Model

The details of the salience model have been described previously (Wilming et al. 2013). Briefly, the bottom-up salience model consists of a weighted linear combination of the set of 48 features described above. Weights were obtained by applying a logistic regression, predicting whether an observation (vector of feature values) was taken from a fixation or from a control location. Control locations were sampled from the spatial bias of fixations, that is, data from the same subject recorded on other

stimuli. The respective feature weights were obtained for each species, observer and stimulus category separately. To evaluate the performance of the saliency model, we performed a leave-one-out cross-validation on the level of observers. This ensured that data from the observer to be predicted were never used during weight estimation. Thus, weights were estimated with data from all other subjects, which allowed us to focus on bottom-up influences that are shared across observers in each species.

Statistical Comparisons

Because we had different numbers of observers across species, we needed to correct for potential statistical differences introduced by this imbalance. We accomplished this in 2 ways. First, we bootstrapped 95% confidence intervals (CIs) for human AUC values by repeatedly selecting 11 random human observers (with replacement) and averaging their AUC values. This estimated the distribution of average AUC values to be expected had only 11 human observers participated in the experiment. Mean monkey AUC values falling outside of the 95% CI for humans were interpreted as support for the hypothesis that monkeys and humans show a true species difference. For visualization purposes, we also bootstrapped 95% CIs for monkey observers by repeatedly sampling 11 monkey observers with replacement. Furthermore, estimates of within-species consistency, measured via AUC, exhibit a dependency on the number of observers used for prediction: fewer observers produce on average smaller AUC values than larger group sizes (Wilming et al. 2011). When comparing humans with monkeys, any difference in within-species consistency is therefore potentially due to different sample sizes. To account for this possibility, we estimated the effect of smaller group sizes in our human data. We sampled groups of observers with $N = 11$ observers 500 times. In each group, we estimated the consistency between human observers. We were also interested in the development of the performance of within-humans consistency over the number of observers used for computing the prediction. We therefore additionally subsampled each group and predicted each observer in the group with different numbers of observers from the same group. We predicted each observer in a group with 1, 2, 3, ..., 10 randomly sampled observers. These subsamplings yielded 500 estimates for each number of predicting observers, which we used to determine intervals that contained 95% of the samples around the estimated mean. Because the estimated mean appeared to follow an exponential function across the number of observers used for prediction, we also fitted an exponential function to the data. This was done to aid data visualization.

Combined Cross-Species and Saliency Model and Model Comparison

We found that the bottom-up saliency prediction and the cross-species prediction gave comparable AUC values for predicting fixation locations of both species. We were therefore interested in evaluating whether the cross-species prediction and the bottom-up saliency model explained unique or shared aspects of the viewing behavior. To investigate this issue, we estimated the predictive power of a combined model, including both predictors in a logistic regression, again classifying image locations as fixated or non-fixated. In detail, to classify fixation and control locations of one human observer, we trained a saliency model on all other human observers and used all monkey fixation data to compute the cross-species prediction. The resulting model therefore contained 2 predictors: bottom-up

saliency and cross-species predictions. Conversely, to classify monkey fixation locations, we used the data of all other monkeys to train the saliency model together with all human fixation data as cross-species prediction (for a schematically depiction of the approach, see Fig. 1D). We predicted each individual observer using a leave-one-observer-out cross-validation. For each logistic regression computed, we standardized the mean and standard deviation of the 2 predictors.

This analysis allowed us to assess whether the combination of bottom-up saliency and the cross-species prediction improves over using either predictor alone. To ensure that a larger number of free parameters in the combined model did not provide a predictive advantage, despite reporting cross-validated test performance, we fitted an additional model in which we combined 2 bottom-up saliency models: one trained on the same-species data and a second one trained on the cross-species data. This joint model has the same number of free parameters as the original joint model (combining bottom-up saliency and cross-species predictions) and therefore allowed us to verify that any observed improvements are due to additionally explained fixation locations, rather than being of pure technical nature. Our rationale for combining the 2 bottom-up saliency models was that the cross-species bottom-up saliency model can only capture those aspects of the cross-species prediction that are driven by bottom-up saliency. Any improvements of the cross-species + saliency model over such a cross-saliency + within-saliency model must therefore be due to factors captured in the cross-species prediction that are not due to bottom-up influences.

Control Experiment: Influence of Head Restraint

We conducted a control experiment to investigate whether the difference in the usage of head restraints across species may explain the results observed. Since the results of this experiment are, strictly speaking, not necessary to evaluate the main results we report them here. The influence of head restraints was accomplished by comparing human viewing behavior with and without head restraints (head-fixed vs. head-free condition). Participants freely viewed all images from the urban and fractals category and additionally carried out a guided viewing task, which required them to fixate a dot on the screen that changed position as soon as the eyes landed on its location.

Participants, Apparatus, and Stimuli

We recruited 19 participants (12 female, 7 male, mean age 24, ranging from 18 to 41). Participants watched all images from the urban and fractals category. Stimuli were displayed on a BenQ XL2420T with a resolution of 1920×1080 ($53.7^\circ \times 30.2^\circ$). Viewing distance was 60 cm. Gaze position was tracked with a remote SR-Research EyeLink 1000 (SR Research Ltd., Ottawa, ON, Canada) with a sampling rate of 500 Hz. During the head-fixed condition, participants bit into a mouth guard that was individually fitted to their teeth. The mouth guard was then attached to a chin rest that was used as an additional restraint. During the head-free condition, chin rest and mouth guard were removed and participants were only instructed to sit still.

Procedure and Task

During the experiment, participants viewed all images in 2 blocks of 64 images that corresponded to the 2 different conditions. Stimuli, condition, and category order were randomized but images from one category were kept as sub-blocks within each condition block. Stimuli and condition order were counter-balanced across subjects such that pairs of participants saw all

128 images. Images were shown for 6 s and were preceded by a cross that had to be fixated before a trial started. After participants viewed all images in a condition block they carried out 10 trials of a guided viewing task. A dot appeared on the screen, which changed position as soon as the participants gaze location was within 2.2° of the dot. Participants had to “chase” the dot, which changed position 200 times. This allowed us to collect data from 2000 saccades per subject per restraint condition. Subjects were instructed to view images freely. The eye tracker was calibrated before each condition block with a 13-point calibration grid (validation error below 0.55°). The entire experiment took about 60–70 min with briefing, fitting of the mouth guard, experiment, breaks, and debriefing.

Analysis and Results

We analyzed the timing of fixations, the distribution of fixation locations, and the saccadic main sequence in both restraint conditions. Fixation durations were analyzed by computing a cumulative histogram for each participant and restraint condition (Supplementary Fig. 1A). Plotting these against each other showed almost perfectly straight lines along the diagonal. To statistically corroborate this, we fitted a regression on a per subject basis and computed the variance explained by the regression line. The line fits had an average slope of 1.01 (SD = 0.05) and provided an exceedingly good fit ($r^2 > 0.99$). Systematic deviations caused by condition differences would curve the relationship between cumulative fixation durations and would therefore not be well described by a linear relationship. The distribution of fixation locations was analyzed by computing within-condition and across-condition AUC prediction scores in analogy to the within-species and cross-species AUC scores. Specifically, we computed how well participants in the head-free and head-fixed condition predicted each other and how well participants from the head-fixed condition predicted the head-free condition and vice versa. We used leave-one-out cross-validation for all AUC scores. This yielded 19 AUC scores per comparison, which we compared by plotting their cumulative histograms. Additionally, we computed paired t-tests which did not reject the null hypothesis of no difference between conditions (Supplementary Fig. 1B and C; urbans: within \leftrightarrow within: $P > 0.4$, cross \leftrightarrow within head fixed: $P > 0.19$, cross \leftrightarrow within head free: $P > 0.84$; fractals: within \leftrightarrow within: $P > 0.21$, cross \leftrightarrow within head fixed: $P > 0.14$, cross \leftrightarrow within head free: $P > 0.38$). We also compared the saccadic main sequence, that is, saccade amplitude versus peak velocity, between head-fixed and head-free conditions. The main sequence in both conditions was highly similar within subject (Supplementary Fig. 1D, $r^2 > 0.98$). In summary, fixing the head in a central position in front of the screen did not change viewing behavior in an appreciable way in this experiment.

Results

We began our comparisons of viewing behavior between species by comparing how well the viewing behavior of individual observers can be predicted by fixation locations from another observer. Each observer in the study, monkey and human, viewed 64 images from 3 different categories (natural, urban, and fractal scenes; example images in Fig. 1A). For each image, we computed the density of fixations across space from one observer (see Fig. 1B for 2 example densities) and used this fixation density as a predictor for fixation locations of another observer. Prediction accuracy was quantified by computing the AUC, separating fixation density values at fixation locations of the predicted observer from density values at fixation locations

on other images viewed by the same observer (“control locations”; see Fig. 1C). The choice of control locations accounts for the influence of the center bias in fixation selection (Tatler et al. 2005; Tatler and Vincent 2009). AUC values around 0.5 indicate chance prediction performance, whereas values close to 1 indicate perfect separation of actual and control fixations. Thus, if the viewing behavior of one observer is similar to the viewing behavior of another, high AUC values can be expected. We computed AUC values for all pairs of observers, yielding 3 similarity matrices (Fig. 2A), one for each stimulus category, that show how well each observer predicted each other observer. We then applied hierarchical agglomerative clustering to these similarity data to test for groups of participants with similar viewing behavior and rearranged the similarity matrices accordingly (see Materials and Methods for more details). This data-driven approach resulted in humans and monkeys sorted into different top-level clusters for natural and urban stimuli (Fig. 2B). The cluster structure for fractals is similar, but 3 human observers are part of the top-level monkey cluster and are then separated at the next level. In all 3 categories, monkeys and humans predicted each other less compared with the same species predicting itself (unpaired t-tests comparing within-species vs. cross-species, all P s < 0.05 with Bonferroni-Holm correction for 6 comparisons; average AUC values; humans = H; monkeys = M; naturals: $H \leftrightarrow M = 0.539$, $H \leftrightarrow H = 0.579$, $M \leftrightarrow M = 0.615$; urbans: $H \leftrightarrow M = 0.589$, $H \leftrightarrow H = 0.699$, $M \leftrightarrow M = 0.626$; fractals: $H \leftrightarrow M = 0.612$, $H \leftrightarrow H = 0.658$, $M \leftrightarrow M = 0.622$). Concentrating only on the monkey clusters, that is, the right branch of the top-level cluster revealed that monkeys were split into 2 groups in all stimulus categories. Interestingly, the same monkeys are assigned to the 2 groups, with the exception of the natural stimuli where one monkey switches clusters. The 2 top-level monkey clusters separate monkeys from the US laboratories (identical rearing conditions; cluster 1) from the macaques at the DPZ, who are, however, joined by 2 other US monkeys (cluster 2) on naturals and urbans and one other monkey for fractals. Differences between monkeys might therefore be in part explained by housing and rearing conditions at the different recording sites. Yet, comparably small numbers of monkeys in the respective clusters do not lend themselves to reliable statistical inference and make it difficult to distinguish natural variation from systematic influences. We therefore do not explore distinctions between monkey clusters any further in this manuscript. Overall, our findings suggest that viewing behavior of monkeys and humans is only comparable to a limited extent, and that viewing behavior across monkeys is not homogeneous.

The clustering analysis raised the question of which aspects of viewing behavior between humans and monkeys are comparable and which are not. To address this question, we investigated how well viewing behavior of both species can be predicted by factors that are known to drive fixation selection strategies in humans. Figure 1D shows a summary of all comparisons performed.

We started by quantifying the similarity of viewing behavior across observers within each species individually (Fig. 1D, blue line). In contrast to the cluster analysis, this “within-species consistency” was computed by predicting individual observers from the data of all other observers. Again using AUC, this allowed us to estimate similarities in viewing behavior within a group, not just behavior shared between pairs of observers (Wilming et al. 2011). At the same time, the within-species consistency is the best-known predictor of individual viewing behavior in humans (Onat et al. 2014). Intuitively, the consistency

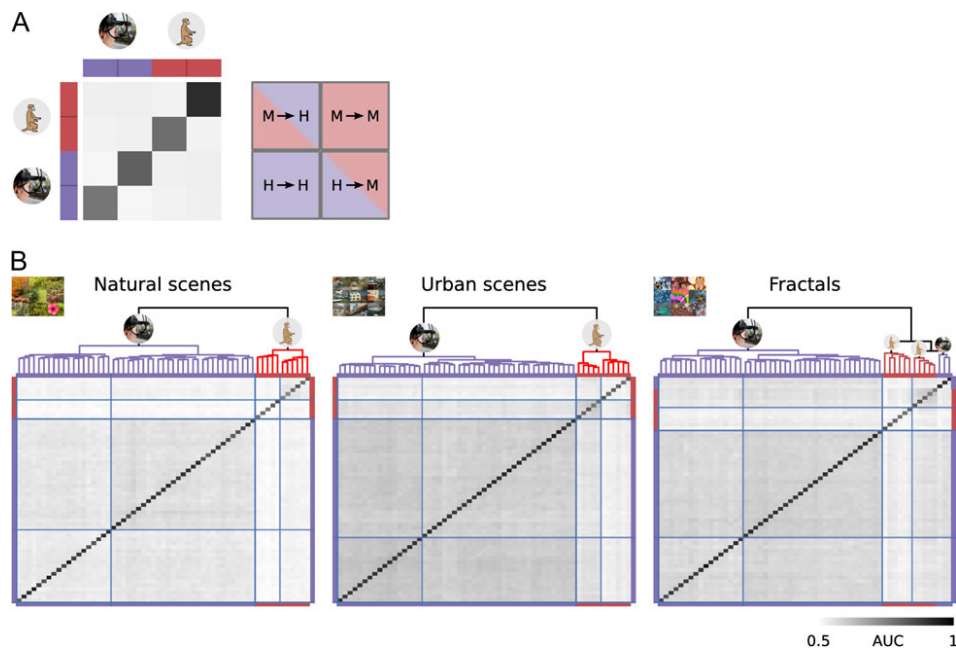


Figure 2. Similarity of viewing behavior between all observers for different categories of stimuli. (A) Schematic of a 4×4 similarity matrix to visualize who predicts who in the full matrix below, for example, H→M depicts areas where human observers predict monkey observers. (B) Full similarity matrix constructed from AUC values between pairs of observers (left = natural scenes, center = urban scenes, right = fractals). The intensity of individual points encode how well an observer predicts another. The species is encoded by different colors on the side of each matrix (purple = humans; red = monkeys). Rows and columns are sorted according to the results of hierarchical clustering. The dendrogram on the top shows the cluster structure, links are colored according to their species composition (purple only humans, red only monkeys, black mixed). Monkeys and humans are sorted into different clusters by the hierarchical clustering algorithm.

is high when all individuals in a group share the same viewing behavior and thus select the same fixation locations. Applied to monkeys and humans, respectively, this analysis revealed that the consistency between humans was higher than between monkeys ($\langle H \rangle_{\text{natural}} = 0.658$, $\langle H \rangle_{\text{urban}} = 0.767$, $\langle H \rangle_{\text{fractal}} = 0.727$, $\langle M \rangle_{\text{natural}} = 0.598$, $\langle M \rangle_{\text{urban}} = 0.680$, $\langle M \rangle_{\text{fractal}} = 0.664$; Fig. 3 top panels, blue symbols, and bars; Table 1 summarizes AUC values numerically; unpaired t-tests $P < 0.05$ with Bonferroni-Holm correction). To exclude the possibility that the estimate of the within-monkey consistency was lower simply because we had less observers available to estimate shared viewing behavior, we estimated the within-human consistency also with fewer observers (Fig. 3, lower right plot in each panel). The difference between within-human and within-monkey consistency persisted even when the number of predicting participants was kept identical in both species, that is, using only 10 human observers for the prediction (Fig. 3, lower right plot in each panel). In particular using groups of 11 humans where each observer is predicted by the remaining 10 yielded consistency estimates that were very close to those when 105 observer predicted 1 observer (mean difference in AUC: 0.002, 0.0001, 0.01). These results demonstrate that for equal sample sizes intra-species predictability is higher for humans than monkeys.

In a next step, we computed how well viewing behavior of one species predicts that of the other species, that is, cross-species consistency (Fig. 1D, red lines). For all observers and images, we computed how well the fixation density of all fixations of the respective other species on the stimulus separated actual and control fixation locations. We then averaged across stimuli and members of the predicted species. This analysis quantifies the amount of shared viewing behavior between species. We found that the cross-species prediction was below the within-species consistency for both species and all categories

(Fig. 3 top panels, red symbols and bars; $\langle M \rightarrow H \rangle_{\text{natural}} = 0.579$, $\langle M \rightarrow H \rangle_{\text{urban}} = 0.666$, $\langle M \rightarrow H \rangle_{\text{fractal}} = 0.672$, $\langle H \rightarrow M \rangle_{\text{natural}} = 0.551$, $\langle H \rightarrow M \rangle_{\text{urban}} = 0.626$, $\langle H \rightarrow M \rangle_{\text{fractal}} = 0.614$; paired t-tests all P s < 0.05 with Bonferroni-Holm correction for 6 tests). In total, the cross-species consistency (M→H) reached 50%, 62%, and 76% of the within-human consistency and 51%, 70%, and 70% of the within-monkey consistency for the 3 stimulus classes, respectively (we subtract the chance level of 0.5 from each AUC value before computing ratios to avoid artificial inflation of these scores). This observation holds not only on average, but for all individual observers, that is, the cross-species is smaller than the within-species consistency for all 11 monkey and all 106 human observers.

Interestingly, we observed that monkeys predicted each other as well as they predicted humans, that is, the cross-species (M→H) score is comparable to the within-monkey score (unpaired t-tests all $P > 0.05$ with Bonferroni-Holm correction for 3 tests; bootstrapped CIs are overlapping; mean differences of AUC values are -0.019 , -0.014 , and 0.008 w.r.t to naturals, urban, and fractals). This is a first indication that the consistent viewing behavior of monkeys is a subset of the consistent viewing behavior of humans. In combination, the cross-species scores and the extent to which monkeys predict humans suggests that both species are driven by similar factors. Furthermore, the larger consistency of human viewing behavior signals the presence of additional factors that are not present in monkeys. The presence of such factors would also explain why the cross-species score is not symmetric.

To better understand similarities and differences across species, we made use of a bottom-up saliency model to test in how far low-level factors could explain the effects observed. We computed a set of 42 different visual features organized on 3 different spatial scales (see Materials and Methods). The resulting saliency model consists of a weighted sum of these features.

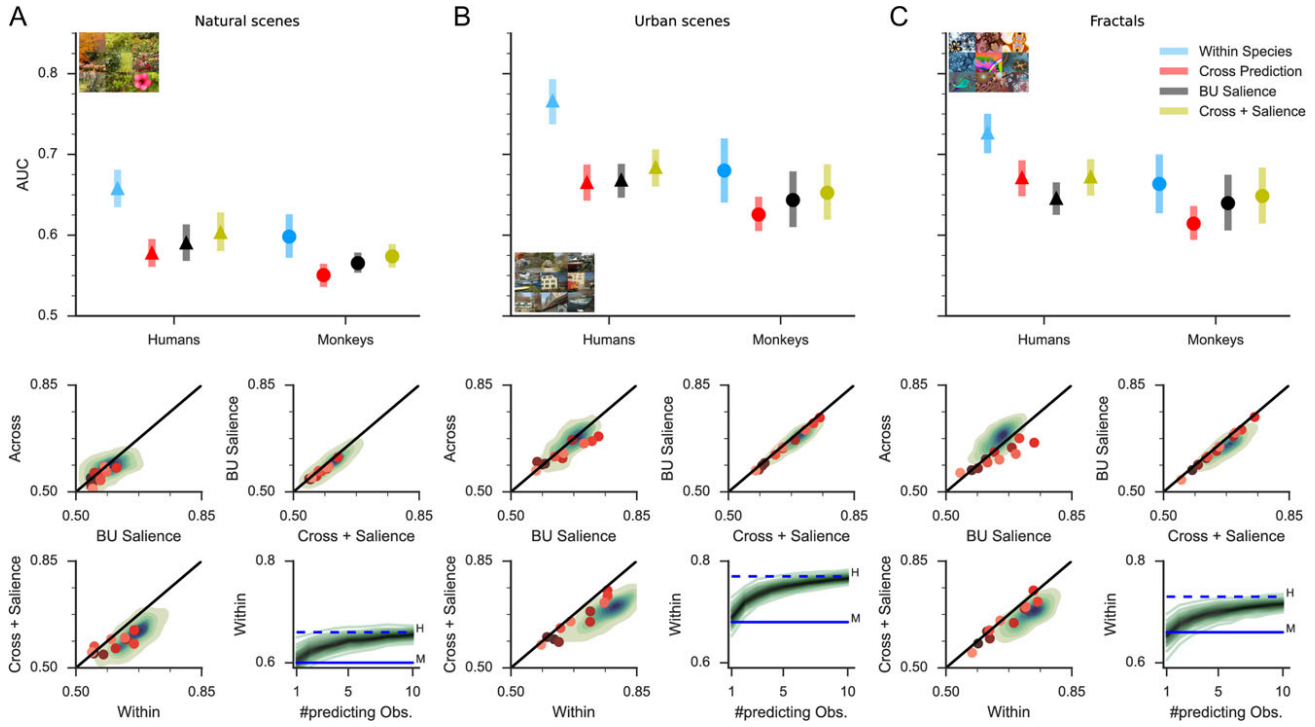


Figure 3. Predicting monkey and human fixation locations with different predictors. Columns show results for natural, urban, and fractal scenes (from left to right). The bar plots on the top show mean value and 95% CIs for individual predictors. CIs are computed by repeatedly sampling 11 observers to allow better comparison between human and monkey data. Three scatter plots at the bottom show individual comparisons. Red dots show individual monkeys (monkeys are indexed by hue). Green contours show a density estimate of the distribution of human AUC; each shade increases the contained amount of observers by 10%. Bottom right plots in each panel show within-human estimates when the number of predicting observers is subsampled. Dashed blue lines indicate our adjusted estimate for the within-human consistency had only 11 observers participated.

Table 1 Summary of AUC scores and percentages relative to the within-species consistency of different predictors

Category	Within		Cross		Saliency		Saliency + cross	
	Human	Monkey	Human	Monkey	Human	Monkey	Human	Monkey
Natural	0.658	0.598	0.579	0.551	0.591	0.565	0.60	0.574
			50%	51%	57%	67%	66%	75%
Urban	0.767	0.680	0.666	0.626	0.669	0.644	0.69	0.653
			62%	70%	63%	80%	69%	85%
Fractal	0.727	0.664	0.672	0.614	0.646	0.640	0.67	0.649
			76%	70%	64%	85%	76%	91%

We chose this type of model because it allowed us to estimate weights for each observer and category independently with a logistic regression that optimally separated actual and control fixations based on their predicted empirical saliency values (rightmost panel in Fig. 1B shows one example saliency map). Furthermore, our model comprised “pure” bottom-up features that do not exploit semantic aspects of the stimuli (e.g. faces, objects, or text). Whereas more complex models can indeed improve predictive performance (as indicated by the success of deep network models, e.g. Kümmerer et al. 2014), they mix lower and higher level features in their prediction. This runs counter to the goal of the current work, which was to separate lower and higher level factors. Our model is therefore designed to act as a probe into bottom-up influences on viewing behavior, not to maximize prediction accuracy. To ensure comparability to the within and cross-species similarity, we use a leave-one-observer-out cross-validation scheme to fit the saliency model. That is, predictions for a specific observer are independent from

the data of that observer. The saliency model therefore must predict behavior of an observer it has not encountered before, which implies that the saliency models can only capitalize on viewing behavior that is consistent between observers. The within-species consistency therefore acts as an upper limit on the performance of the saliency model (Wilming et al. 2011). Once fitted, the saliency model was applied in the same manner as the within- and cross-species predictions (Fig. 1D, black line). Correspondingly, the model prediction accuracy was computed based on AUC values. Figure 3 shows the results of the bottom-up saliency model. Black triangles and circles show the respective mean performance for humans and monkeys ($\langle H \rangle_{\text{natural}} = 0.591$, $\langle H \rangle_{\text{urban}} = 0.669$, $\langle H \rangle_{\text{fractal}} = 0.646$, $\langle M \rangle_{\text{natural}} = 0.565$, $\langle M \rangle_{\text{urban}} = 0.644$, $\langle M \rangle_{\text{fractal}} = 0.640$). CIs for human and monkey bottom-up saliency AUCs are largely overlapping (Fig. 3, black bars, unpaired t-tests all $P > 0.05$ with Bonferroni-Holm correction). To further investigate the similarity of the bottom-up saliency model in both species, we compared the predictive

power of individual features that enter the saliency model by computing feature-specific AUC values. We found that the resulting feature-fixation AUCs were highly correlated across species on urban scenes and fractals and to some extent on natural scenes (Fig. 4). A linear regression between feature AUC values of humans and monkeys showed that the pattern of human feature-fixation AUC values explained 94% and 97% of the variance between the monkey feature-fixation AUC values for urbans and fractals and 27% on natural scenes. To estimate the distribution of regression slopes, we recalculated the linear regression on randomly sampled subsets of 11 human and monkey observers (sampling with replacement, see Materials and Methods for more details). The distributions of slopes, across bootstrapping samples, on urban and fractal scenes were approximately centered on 1 for urban and fractal scenes but 1 was not contained in the 95% interval for natural scenes (average slopes are: 0.52, 0.90, and 0.97 for naturals, urbans, and fractal scenes, respectively). These results show that feature-fixation AUCs were almost identical for urban and fractal scenes and thus the models were nearly identical for these stimulus categories. The finding of smaller feature AUC values on naturals is in line with the observation that, in general, all AUC values are lower on natural scenes compared with urban and fractal scenes. Yet, the pattern of results within this category is highly similar to the other 2 categories (see Fig. 3A). This suggests that the observed differences are not related to species differences but rather to an overall category difference. In summary, the results of the bottom-up saliency models and similarities in feature AUCs suggest that viewing behavior that is shared between species is to a large extent driven by bottom-up saliency.

Interestingly, bottom-up saliency reached 67%, 80%, and 85% of the within-monkey consistency AUC scores, suggesting that bottom-up saliency explains a large part of the consistent viewing behavior between monkeys. In humans, however, we observed only 57%, 63%, and 64%, indicating that human viewing behavior is strongly guided by factors beyond low-level features (unpaired t-tests, all $P < 0.05$ with Bonferroni-Holm correction for 3 tests). These numbers imply that, across categories, 23% of the consistent viewing behavior in monkeys cannot be explained by bottom-up saliency while the gap is 39% for humans. This demonstrates that although the bottom-up saliency models of both species are similar and explain viewing behavior to a comparable degree, the gap to the respective within-species consistency leaves a large fraction of human viewing behavior to be explained.

With this in mind, it is interesting to know whether there are other factors contributing to successful cross-species predictions, besides a presumably shared bottom-up selection mechanisms. Despite the fact that bottom-up saliency AUC values were larger than the cross-species prediction for the majority of monkeys (8/11, 8/11, 10/11 w.r.t. naturals, urbans, and fractals), both factors might explain different aspects of viewing behavior. To explicitly investigate this possibility, we tested one additional linear model that combined bottom-up saliency and the cross-species prediction (Fig. 1D, yellow lines). We reasoned that performance of the combined model could only improve over the individual predictors, if both predictors explain, at least in part, different aspects of the consistent viewing behavior. The combined model used a logistic regression to obtain optimal weights for z-scored bottom-up saliency values (with the weighting of the individual features optimized

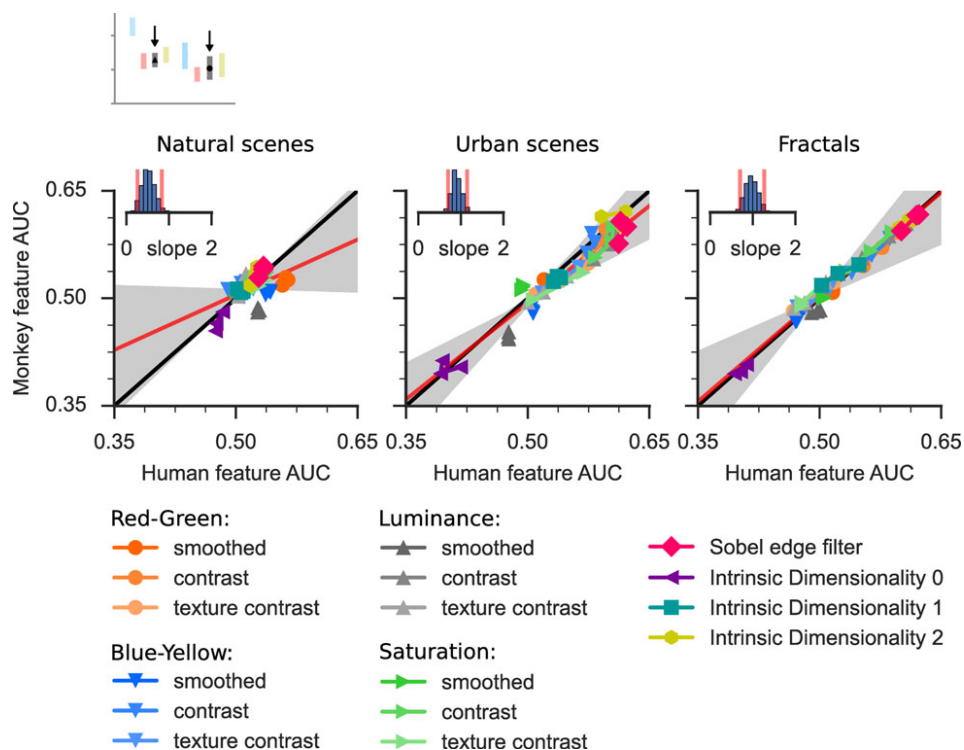


Figure 4. Comparison of bottom-up saliency across species. Panels show feature-fixation AUC values for humans and monkeys (see Materials and Methods for feature definitions). Red lines show the best fit linear model that explains monkey feature-fixation AUCs based on human feature-fixation AUCs. The shaded blue area contains 95% of all bootstrapped regression lines created by repeatedly subsampling human and monkey observers with replacement. Small insets show the bootstrapped distribution of slopes of the linear regression. The area between the red bars contains 95% of bootstrapped slopes. Columns show different categories (naturals, urbans, and fractals).

as before, that is, by leave-one-observer-out cross-validation), and z-scored cross-species predictions. Figure 5 plots the bottom-up saliency values against cross-species predictions for fixated locations and non-fixated locations. Model predictions can best be understood in terms of a classifier that compares the respective bottom-up saliency and cross-species values of a test fixation against a given threshold (here zero for both models). Fixations with values exceeding the threshold are classified as “fixated”, all others as “non-fixated” (thresholds shown in Fig. 5 as dashed black lines; red coloring indicates more actuals than control fixations, blue the reverse). As can be seen in Fig. 5, a model combining low-level saliency and cross-species predictions can only improve over the 2 individual models when a combination of both models, that is, a slanted decision function, classifies more points correctly (i.e. better separates red from blue points). We found that the combined model exhibited only small, but consistent, improvements over bottom-up saliency and the cross-species prediction alone, for all monkeys on all categories (paired t-tests all $P < 0.05$ with Bonferroni-Holm correction for 6 tests; differences of combined-saliency AUC values, naturals: 0.008, urbans: 0.009, fractals: 0.009). The facts that the decision function of the combined

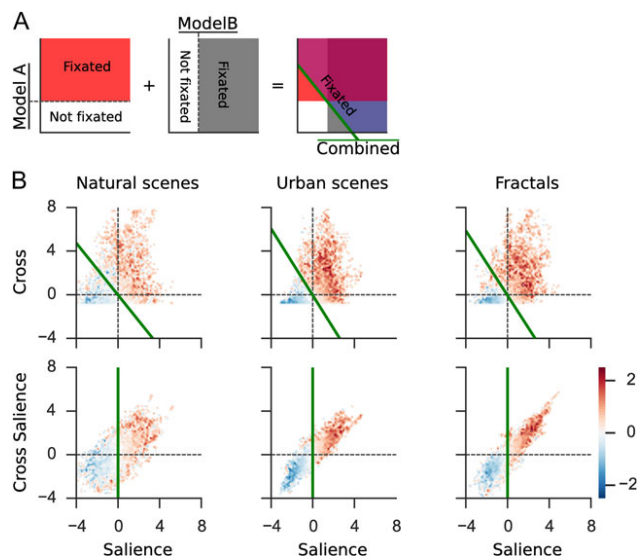


Figure 5. Combination of different models to predict monkey fixation locations. (A) Cartoon drawing to exemplify how 2 models form a combined model. Model A and B both project data into a space in which samples are classified as fixated or non-fixated based on their position w.r.t a threshold (dashed lines). The combined model weights and sums both models to potentially improve its prediction. The slope of the resulting decision function (green line) depicts the relative impact of the 2 combined models. (B) Performance of the combined model on different stimulus categories (naturals, urbans, fractals). Color encodes the log likelihood of observing fixations with a specific combination of predictor values in the data set (e.g. $\log(P(\text{fix} | \text{cross prediction, within saliency})/P(\text{control} | \text{cross prediction, within saliency}))$ in the top row and $\log(P(\text{fix} | \text{cross-saliency, within saliency})/P(\text{control} | \text{cross-saliency, within saliency}))$ in the bottom row), that is, redish colors imply more actual fixations and blueish colors non-fixated locations. Top row: Each panel shows the separating decision function of a within-species bottom-up model and the cross-species prediction (vertical and horizontal dashed lines, respectively). Diagonal green lines show the decision function of the model that combines within-species saliency and the cross-species prediction. The bottom row shows decision functions for within-species saliency and cross-species saliency models. Here, the decision function (green line) for a model that combines within-species and cross-species saliency falls onto the function for the within-species saliency model, that is, the combined model utilizes exclusively the species-specific saliency model, neglecting the respective model derived from the other species.

model is slanted (Fig. 5B, green lines), and that the combined model improves over its constituent models, together imply that bottom-up saliency and the cross-species prediction both significantly contribute to the prediction. Yet, it has to be considered that the dynamic range, that is, the difference between the bottom-up saliency model and the intra-species predictability is rather small. As a consequence, the performance of the combined model was very close to the within-monkey estimate (naturals: 0.574 vs. 0.598; urbans: 0.653 vs. 0.68, fractals: 0.649 vs. 0.664; 75%, 85%, and 91%, respectively). For humans, the combined model showed marginal improvements over either bottom-up saliency or the cross-species prediction (naturals: 0.01, paired t-tests against bottom-up saliency $P < 0.05$; urbans: 0.02, paired t-tests against bottom-up saliency $P < 0.05$; fractals: 0.001, paired t-tests against bottom-up saliency $P < 0.05$; Bonferroni-Holm correction for 6 tests) but was still far away from the within-human estimate (naturals: 0.60 vs. 0.66, urbans: 0.69 vs. 0.77, fractals: 0.67 vs. 0.73; 66%, 69%, 76% relative to the within-human estimate). For both species, these results demonstrate that shared factors, beyond bottom-up selection, are very limited. Rather, shared viewing behavior is predominantly driven by low-level selection.

Although very little improvement was observed with the combined model, with respect to the question whether or not monkeys are a good model system for human fixation behavior, it is nevertheless interesting to investigate why the combined cross-species + saliency predicts monkey fixations better than bottom-up saliency alone. In particular, we wanted to test whether the human-cross-species prediction contains explanatory power beyond its bottom-up saliency component. As a reference, we compared the combined model (saliency + cross-species) with a model that combines 2 saliency models: a cross-species (trained on all human observers) and a within-species model (trained on leave-one-observer-out data from monkey observers), as described earlier. We found that the combined-saliency model did not use the cross-species saliency values for its prediction. This implies that the combined-saliency model is identical to the one obtained by using within-monkey saliency only (the decision function of the combined-saliency model falls onto the decision function of the within-monkey saliency model in Fig. 5; put differently, the weight assigned to the cross-species bottom-up saliency model by the logistic regression is close to zero). This suggests that the little improvement of the combination of the cross-species prediction and bottom-up saliency over bottom-up saliency alone is due to factors that are contained in human viewing behavior but not in our bottom-up saliency model.

Taken together, bottom-up saliency is able to account for most of the consistent viewing behavior between monkeys, while human viewing behavior makes only a small contribution when predicting monkey behavior. For humans, too, the addition of monkey data adds little to improve predictive performance beyond a pure bottom-up model. Moreover, while consistent macaque viewing behavior is well understood in terms of low-level saliency, human viewing behavior is more complex. Thus, our data indicate that both species share similar low-level selection mechanisms, whereas similarities beyond low-level are very limited.

Discussion

The present data revealed that humans and monkeys form distinct clusters of viewing behavior, where predictions within each species are significantly better than predictions across

species. Within species, we found that patterns of human fixation locations are more consistent than those of monkeys, even when we account for the low number of monkey observers. This implies that humans and monkeys do not use identical selection strategies in a free-viewing task; a task that requires no training for monkeys or humans. Despite these overall differences, prediction accuracies across species are significantly above-chance level. This implies that monkeys and humans do share some degree of fixation selection strategies. To investigate in what aspects the selection strategies of these species overlap we compared the predictive power of a set of low-level stimulus features. This revealed remarkable similarities across species. First, the overall predictive power of low-level saliency models is comparable. Additionally, the predictive power of individual low-level features was strongly correlated between species, indicating that similar low-level selection mechanisms are at work in both species. Following this, we compared whether the fixation data of one species adds explanatory power beyond such low-level factors for the respective other species. Here we found only small improvements in predictive power, indicating that, at least in the current data set, low-level saliency is the most important shared component in the guidance of eye movements across species. Taken together, our findings show that free viewing of pictures produces consistent behavior in monkeys that can be related to human behavior in a meaningful way. The large influence of bottom-up saliency in both monkeys and humans, the large number of data points that can be acquired in a short period of time and the fact that no training is required, make free viewing in monkeys very well suited to the study of the neural basis of low-level stimulus-driven oculomotor control. Yet, our results reveal important limitations, as viewing commonalities were limited to low-level selection, excluding shared higher level selection mechanisms. They, therefore, have large implications for future electrophysiological studies in macaques and behavioral comparisons across species.

Although the current study is the largest to date, it is not the first to compare viewing behavior in macaques and humans. Consistent with our findings, other groups, too, reported above-chance predictions of fixation points by saliency models in both humans and monkeys for videos clips (Berg et al. 2009) and gray scale images (Einhäuser et al. 2006). However, there is disagreement in how far the effects are qualitatively comparable in humans and monkeys. Einhäuser et al. (2006) argue that both species are equally driven by bottom-up saliency, while Berg et al. (2009) report that bottom-up saliency is more predictive for human eye-movement behavior. Using a considerably larger set of observers and colored stimuli from 3 different categories, we do not find consistent differences across species that would argue for a true species differences. It should be noted that the low-level AUC values reported previously (Einhäuser et al. 2006; Berg et al. 2009) indicate only a small influence of stimulus features ($AUC \leq 0.59$) in monkeys during free viewing. This is in line with the current monkey data for natural scenes. However, looking at urbans and fractals, we observe larger AUC values for a bottom-up saliency model ($AUC = 0.65$ and 0.64). Bottom-up saliency could therefore be more prominent in guiding monkey eye movements than previously thought, dependent on the stimulus category used. Together with the large correlations in the feature AUCs across species, this data suggest comparable low-level selection mechanisms in macaques and humans. Moreover, while we only observe very small similarities beyond low-level stimulus features in the current free-viewing paradigm, it is possible that other experimental settings or

categories reveal larger consistencies. For instance, similarities between monkeys and humans exist during the viewing of faces (Guo et al. 2003, 2009; Ghazanfar et al. 2006; Shepherd et al. 2010) and scenes containing simple social interactions (McFarland et al. 2013; Solyst and Buffalo 2014). It should be noted, however, that these studies did not estimate the contribution of low-level saliency. It therefore remains unclear in how far the shared viewing behavior observed was driven by shared low-level selection mechanisms.

The interpretation of our results depends on how similar recording conditions were between species. Naturally, there are a few differences between humans and monkeys that we were not able to control. These include the fact that monkeys needed to be head fixed during recordings, that monkeys had less exposure to the kinds of stimuli that were presented, and that monkeys received reward during calibration trials and, more generally, underwent extensive training regimes for other tasks. We will discuss these issues in turn in the next paragraphs.

First, all of the monkeys were head fixed during the recordings. The human participants, however, were only instructed not to move their head, but were otherwise not constrained. To rule out the possibility that a head restraint interferes with fixation selection, we conducted a control experiment in which participants viewed urban scenes and fractals, with and without a head restraint (custom molded bite-bar and chin rest). We analyzed 3 aspects of fixation behavior: fixation durations, selected fixation locations, and the saccadic main sequence. Fixation durations were almost identical between conditions (linear fits had an average slope of 1.01 with $SD = 0.05$, $r^2 > 0.99$, Supplementary Fig. 1). We compared fixated locations across conditions by computing within-condition and across-condition AUC prediction scores, in analogy to the within-species and cross-species AUC scores. The resulting AUC values were very similar (Supplementary Fig. 1B and C; $P > 0.21$ for all comparisons), again indicating no systematic differences across conditions. Finally, no significant differences were observed between the saccadic main sequence between species in both conditions (within subject $r^2 > 0.98$). In summary, fixing the head in a central position in front of the screen did not change the overall pattern of viewing behavior. We therefore conclude that the difference in head restraint between species is not relevant for the interpretation of our results.

Second, compared with humans, monkeys have limited prior exposure to the kind of stimuli shown in this study (urban and natural stimulus categories). Thus, differences in specific associations, memories, and emotions, triggered for humans but not for macaques, might explain why we find consistent viewing behavior beyond bottom-up saliency in humans but not in monkeys. What speaks against this possibility is the fact that we observed a very similar pattern of results across all categories, including fractal stimuli, for which both species have little to no prior exposure. Accordingly, our results with fractal stimuli corroborate our conclusions even for previously unknown stimulus categories.

Third, to achieve successful calibration, monkeys needed either to detect a color change of a rectangle located at random screen locations or fixate a cross at random screen locations. More generally, monkeys were previously trained in a wide variety of tasks that associated specific actions with rewards (DPZ monkeys: classification of 2D and 3D random dot stimuli, grasping tasks, and fixation task; YNPRC and WNPRC: delayed match to sample, change detection, visual search, and covert attention task). Is it therefore conceivable that monkeys searched the free-viewing stimuli for reward? We believe that

this is unlikely for several reasons. First, monkey fixation durations on free-viewing images are much shorter than the minimum fixation requirements during reward trials (color change task: 500–1100 ms required, average fixation duration: 225 ± 153 ms; fixation task: 1250 ms required, average fixation duration: 257 ± 214 ms). This shows that the monkeys did not simply transfer the temporal properties of reward providing actions to the free-viewing trials. Second, if operant conditioning engendered specific consistent viewing strategies, we should have observed elevated within-monkey AUC scores. In contrast, within-monkey AUC scores did not substantially exceed the scores for bottom-up saliency. Finally, speaking specifically against an influence of interleaved calibration, the fixation epochs during calibration trials are in close analogy to drift correction procedures that preceded stimulus presentation in the human experiments. Forced fixation epochs were therefore present in both human and monkey experiments. Hence, we believe that neither training for other tasks nor calibration procedures undermine the interpretation of our results.

Our results suggest an effective, and comparable saliency model in both species. However, whether or not bottom-up saliency has a causal influence on eye movements is a matter of ongoing debate (Li 2002; Einhäuser and König 2003; Mazer and Gallant 2003; Henderson et al. 2007; Einhäuser et al. 2008; Arcizet et al. 2011; Schütz et al. 2011; Betz et al. 2013). In many cases, saliency models have been used successfully to predict eye-movement targets during free viewing of images (Itti and Baldi 2005; Kienzle et al. 2007; Zhang et al. 2008; Bruce and Tsotsos 2009; Hwang et al. 2009; Judd et al. 2009; Zhao and Koch 2011). Furthermore, recent reports provided evidence for the existence of a functional saliency map in the human brain (Bogler et al. 2011; Ossandón et al. 2012). In summary, there is good evidence for the existence of saliency-like selection mechanisms in the brain. However, even if saliency turns out to be a correlate, for example, of object detection (Einhäuser et al. 2008; Nuthmann and Henderson 2010), rather than a true selection mechanism, the current finding of above-chance cross-species predictions, and comparable low-level AUC values still indicate shared mechanisms in the guidance of overt visual attention.

The results that we have observed are on a purely behavioral level, and it therefore remains an open question as to how these map onto the neural level. On the one hand, it is known that large homologies exist between human and monkey early visual (Orban et al. 2004) and higher level ventral visual areas (Kriegeskorte et al. 2008; Kornblith et al. 2013; Yovel and Freiwald 2013; Cichy et al. 2014). It is therefore tempting to hypothesize that the observed similarities in viewing behavior are the result of similar neural processing of visual space. On the other hand, marked differences exist across species. The temporal lobe is much larger in humans (Rilling and Seligman 2002) and at the same time, fewer and less clear homologous brain structures exist in dorsal areas (Orban et al. 2004). For example, human LIP possesses more retinotopically organized areas than monkey LIP (Patel et al. 2010). And, on a larger scale, oculomotor control in the human brain appears to be more lateralized while the monkey brain shows more contralateral specificity (Kagan et al. 2010; Oleksiak et al. 2011). It is therefore possible that the human brain possesses different mechanisms to drive eye-movement behavior. In this context, our behavioral findings also resonate well with the absence of a ventral attention network in macaque monkeys but overlap of the dorsal attention network in both species during a paradigm that required detection of target images in a rapidly presented stream of images (Patel et al. 2015). Many areas in the dorsal

attention network are retinotopically organized and the network likely contains a priority map for guiding eye movements (Bisley 2011), making it a plausible candidate for stimulus-driven control of eye movements. Thus, while the current set of results suggests similar low-level selection mechanisms, our analyses additionally suggest that homologies w.r.t. higher level selection of fixation locations are limited and cannot be assumed a priori for electrophysiological studies on macaques. Comparisons of the neural basis of higher level oculomotor selection mechanisms across species should therefore be validated by behavioral comparisons—ideally with tasks that require the same amount of training for both species. Finally, we would like to emphasize that our results do not preclude similarities between species during other cognitive tasks (e.g. Wisconsin Card Sorting Task (Nakahara et al. 2002) or covert attention tasks (Caspari et al. 2015)).

Supplementary Material

Supplementary material are available at *Cerebral Cortex* online.

Funding

The authors gratefully acknowledge the funding of EU Grants ERC-2010-AdG-269716 “Multisense” and H2020-FeT-PROaCT-2014 641321—“socSMCs”. This work was supported by National Institute of Mental Health Grants MH080007 and MH093807 (to E.A.B.) and the National Institutes of Health, ORIP-OD010425. C.X. and S.T. were supported by the Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center 889 “Cellular mechanisms of sensory processing” and the Research Unit 1847 “The physiology of distributed computing underlying higher brain functions in non-human primates” and a DFG Postdoctoral Fellowship to T.C. Kietzmann.

Notes

We would like to thank Alexandra Vormberg for assistance with data collection for the head restraint experiment. *Conflict of Interest:* None declared.

References

- Açık A, Sarwary A, Schultze-Kraft R, Onat S, König P. 2010. Developmental changes in natural viewing behavior: bottom-up and top-down differences between children, young adults and older adults. *Front Psychol.* 1:207.
- Arcizet F, Mirpour K, Bisley JW. 2011. A pure saliency response in posterior parietal cortex. *Cereb Cortex.* 21:2498–2506.
- Berg DJ, Boehnke SE, Marino RA, Munoz DP, Itti L. 2009. Free viewing of dynamic stimuli by humans and monkeys. *J Vis.* 9:1–15.
- Betz T, Kietzmann T, Wilming N, König P. 2010. Investigating task-dependent top-down effects on overt visual attention. *J Vis.* 10:1–14.
- Betz T, Wilming N, Bogler C, Haynes J-D, König P. 2013. Dissociation between saliency signals and activity in early visual cortex. *J Vis.* 13:1–12.
- Bisley JW. 2011. The neural basis of visual attention. *J Physiol.* 589:49–57.
- Bogler C, Bode S, Haynes J-D. 2011. Decoding successive computational stages of saliency processing. *Curr Biol.* 1–5.
- Bruce NDB, Tsotsos J. 2009. Saliency, attention, and visual search: an information theoretic approach. *J Vis.* 9:1–24.

- Calapai A, Berger M, Niessing M, Heisig K, Brockhausen R, Treue S, Gail A. 2016. A cage-based training, cognitive testing and enrichment system optimized for rhesus macaques in neuroscience research. *Behav Res Methods*. doi:10.3758/s13428-016-0707-3.
- Caspari XN, Janssens T, Mantini XD, Vandenberghe XR, Vanduffel XW. 2015. Covert shifts of spatial attention in the macaque monkey. *J Neurosci*. 35:7695–7714.
- Castelhano M, Mack M, Henderson J. 2009. Viewing task influences eye movement control during active scene perception. *J Vis*. 9:1–15.
- Cichy RM, Pantazis D, Oliva A. 2014. Resolving human object recognition in space and time. *Nat Neurosci*. 17:1–10.
- Einhäuser W, König P. 2003. Does luminance-contrast contribute to a saliency map for overt visual attention? *Eur J Neurosci*. 17:1089–1097.
- Einhäuser W, Kruse W, Hoffmann K-P, König P. 2006. Differences of monkey and human overt attention under natural conditions. *Vision Res*. 46:1194–1209.
- Einhäuser W, Spain M, Perona P. 2008. Objects predict fixations better than early saliency. *J Vis*. 8:1–26.
- Ghazanfar AA, Nielsen K, Logothetis NK. 2006. Eye movements of monkey observers viewing vocalizing conspecifics. *Cognition*. 101:515–529.
- Guo K, Meints K, Hall C, Hall S, Mills D. 2009. Left gaze bias in humans, rhesus monkeys and domestic dogs. *Anim Cogn*. 12:409–418.
- Guo K, Robertson RG, Mahmoodi S, Tadmor Y, Young MP. 2003. How do monkeys view faces? A study of eye movements. *Exp brain Res*. 150:363–374.
- Henderson J, Brockmole J, Castelhano M, Mack M. 2007. Visual saliency does not account for eye movements during visual search in real-world scenes. In: Roger PG Van Gompel, Martin H. Fischer, Wayne S. Murray, Robin L. Hill, editors. *Eye movements: A Window on Mind and Brain*. Amsterdam: Elsevier Science. p. 537–562.
- Hwang AD, Higgins EC, Pomplun M. 2009. A model of top-down attentional control during visual search in complex scenes. *J Vis*. 9:1–18.
- Itti L, Baldi P. 2005. Bayesian surprise attracts human attention. *Vision Res*. 49:1295–1306.
- Itti L, Koch C. 2001. Computational modelling of visual attention. *Nat Rev Neurosci*. 2:194–203.
- Johnson L, Sullivan B, Hayhoe M, Ballard D. 2014. Predicting human visuomotor behaviour in a driving task. *Phil Trans R Soc B*. 369:20130044.
- Judd T, Ehinger K, Durand F, Torralba A. 2009. Learning to predict where humans look. *Comput Vis*. 2106–2113.
- Jutras MJ, Buffalo EA. 2010. Recognition memory signals in the macaque hippocampus. *Proc Natl Acad Sci USA*. 107:401–406.
- Jutras MJ, Fries P, Buffalo EA. 2009. Gamma-band synchronization in the macaque hippocampus and memory formation. *J Neurosci*. 29:12521–12531.
- Kagan I, Iyer A, Lindner A, Andersen RA. 2010. Space representation for eye movements is more contralateral in monkeys than in humans. *Proc Natl Acad Sci USA*. 107:7933–7938.
- Kienzle W, Wichmann F, Scholkopf B. 2007. A nonparametric approach to bottom-up visual saliency. *Adv Neural Inf Process Syst*. 19:689–696.
- Kietzmann T, Geuter S, König P. 2011. Overt visual attention as a causal factor of perceptual awareness. *PLoS One*. 6(7):e22614.
- Kietzmann T, König P. 2015. Effects of contextual information and stimulus ambiguity on overt visual sampling behavior. *Vision Res*. 110:76–86.
- Kollmorgen S, Nortmann N, Schröder S, König P. 2010. Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. *PLoS Comput Biol*. 6: e1000791.
- Kornblith S, Cheng X, Ohayon S, Tsao DY. 2013. A network for scene processing in the macaque temporal lobe. *Neuron*. 79: 766–781.
- Kriegeskorte N, Mur M, Ruff D, Kiani R. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*. 60:1126–1141.
- Kümmerer M, Theis L, Bethge M. 2014. Deep gaze I: boosting saliency prediction with feature maps trained on imageNet. *arXiv*. 1411.1045.
- Land MF, Hayhoe M. 2001. In what ways do eye movements contribute to everyday activities? *Vision Res*. 41:3559–3565.
- Land MF, Tatler BW. 2001. Steering with the head: the visual strategy of a racing driver. *Curr Biol*. 11:1215–1220.
- Li Z. 2002. A saliency map in primary visual cortex. *Trends Cogn Sci*. 6:9–16.
- Mazer J a, Gallant JL. 2003. Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron*. 40:1241–1250.
- McFarland R, Roebuck H, Yan Y, Majolo B, Li W, Guo K. 2013. Social interactions through the eyes of macaques and humans. *PLoS One*. 8:e56437.
- Nakahara K, Hayashi T, Konishi S, Miyashita Y. 2002. Functional MRI of macaque monkeys performing a cognitive set-shifting task. *Science*. 295:1532–1536.
- Nuthmann A, Henderson J. 2010. Object-based attentional selection in scene viewing. *J Vis*. 10:1–19.
- Oleksiak A, Postma A, van der Ham IJ, Klink PC, van Wezel RJ. 2011. A review of lateralization of spatial functioning in nonhuman primates. *Brain Res Rev*. 67:56–72.
- Onat S, Açıık A, Schumann F, König P. 2014. The contributions of image content and behavioral relevancy to overt attention. *PLoS One*. 9:e93254.
- Orban G a, Van Essen D, Vanduffel W. 2004. Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn Sci*. 8:315–324.
- Ossandón JP, Onat S, Cazzoli D, Nyffeler T, Müri R, König P. 2012. Unmasking the contribution of low-level features to the guidance of attention. *Neuropsychologia*. 50: 3478–3487.
- Parkhurst D, Law K, Niebur E. 2002. Modeling the role of saliency in the allocation of overt visual attention. *Vision Res*. 42:107–123.
- Patel GH, Shulman GL, Baker JT, Akbudak E, Snyder AZ, Snyder LH, Corbetta M. 2010. Topographic organization of macaque area LIP. *Proc Natl Acad Sci USA*. 107:4728–4733.
- Patel GH, Yang D, Jamerson EC, Snyder LH, Corbetta M, Ferrera VP. 2015. Functional evolution of new and expanded attention networks in humans. *Proc Natl Acad Sci*. 112: E5377–E5377.
- Petersen SE, Posner MI. 2012. The attention system of the human brain: 20 years after. *Annu Rev Neurosci*. 35:73–89.
- Rilling JK, Seligman RA. 2002. A quantitative morphometric comparative analysis of the primate temporal lobe. *J Hum Evol*. 42:505–533.
- Schütz AC, Braun DI, Gegenfurtner KR. 2011. Eye movements and perception: a selective review. *J Vis*. 11:1–30.
- Shepherd S V, Steckenfinger SA, Hasson U, Ghazanfar AA. 2010. Human-macaw gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr Biol*. 20: 649–656.

- Smith TJ, Henderson JM. 2011. Does oculomotor inhibition of return influence fixation probability during scene search? *Atten Percept Psychophys.* 73:2384–2398.
- Solyst JA, Buffalo EA. 2014. Social relevance drives viewing behavior independent of low-level salience in rhesus macaques. *Front Neurosci.* 8:1–13.
- Sullivan B, Johnson L, Rothkopf C, Ballard D, Hayhoe M. 2012. The role of uncertainty and reward on eye movements in a virtual driving task. *J Vis.* 12:1–17.
- Tatler B. 2007. The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *J Vis.* 7: 1–17.
- Tatler BW, Baddeley RJ, Gilchrist ID. 2005. Visual correlates of fixation selection: effects of scale and time. *Vision Res.* 45: 643–659.
- Tatler BW, Vincent BT. 2009. The prominence of behavioural biases in eye guidance. *Vis cogn.* 17:1029–1054.
- Torralba A, Oliva A, Castelano MS, Henderson JM. 2006. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol Rev.* 113:766–786.
- Ward JH. 1963. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc.* 58:236–244.
- Wilming N, Betz T, Kietzmann TC, König P. 2011. Measures and limits of models of fixation selection. *PLoS One.* 6:e24038.
- Wilming N, Harst S, Schmidt N, König P. 2013. Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS Comput Biol.* 9:e1002871.
- Yovel G, Freiwald WA. 2013. Face recognition systems in monkey and human: are they the same thing? *F1000Prime Rep.* 5:10.
- Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW. 2008. SUN: a Bayesian framework for saliency using natural statistics. *J Vis.* 8:32.
- Zhao Q, Koch C. 2011. Learning a saliency map using fixated locations in natural scenes. *J Vis.* 11:1–15.

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Does Correlated Firing Underlie Attention Deployment in Frontal Cortex?

Moein Esghaei^{1,2*} and Cheng Xue^{2*}

¹Cognitive Neurobiology Laboratory, School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran 19395, Iran, and

²Cognitive Neuroscience Laboratory, German Primate Center (DPZ), Goettingen 37077, Germany

Review of Oemisch et al.

Attention is an essential cognitive ability that animals and humans rely on to survive. It allows complex nervous systems to selectively process the most behaviorally relevant sensory information. While an abundance of literature has demonstrated attentional modulation of neuronal activity in sensory areas (Treue, 2001; Maunsell and Treue, 2006), it is much less clear which brain areas control this modulation (Petersen and Posner, 2012). One candidate region is the prefrontal cortex (PFC), which makes reciprocal projections with almost all sensory and motor cortical areas, as well as many subcortical structures (Miller and Cohen, 2001). Studies on the cross-areal interaction between prefrontal and sensory areas have identified PFC as a major control center for directing attention to a location (Moore and Armstrong, 2003; Gregoriou et al., 2009), a feature, or an object (Zaksas and Pasternak, 2006; Baldauf and Desimone, 2014). An additional candidate is anterior cingulate cortex (ACC), which has a close functional connectivity with PFC (Womelsdorf et al., 2014). Although numerous studies have

probed the role of these two frontal areas in controlling attention, it is not clear how attention signals are integrated within and between these regions.

In a recent paper, Oemisch et al. (2015) studied correlations between instantaneous firing rates of neurons located in different frontal cortex substructures in an attention task. They trained two rhesus monkeys to maintain their gaze on a central fixation point on a computer screen while two peripheral drifting gratings, each with a different color, were presented to the left and right visual hemifield. After the color of the fixation point changed to that of either of the gratings, the monkeys had to covertly attend to the corresponding grating (the target) without diverting their gaze from the fixation point (for experimental details, see Fig. 1). The authors recorded single and multi-unit activities from dorsal and lateral PFC, ventromedial PFC, and ACC while the monkeys performed the task.

Oemisch and colleagues (2015) report three major findings. First, starting 280 ms after the monkeys were cued to the target location, the instantaneous activities of those pairs of neurons that were recorded simultaneously became correlated across trials. These correlations were especially pronounced for cross-areal neuron pairs recorded from ACC and PFC, suggesting that a neural interaction occurs between these two areas. This correlation does not appear to be a simple side effect of engaging spatial attention,

but rather appears to reflect where the monkeys attend, because a significant portion of neuron pairs from ACC and PFC showed distinct correlations when the monkeys attended contralaterally rather than ipsilaterally to the recorded hemisphere (Oemisch et al., 2015, their Fig. 5Biii). Second, the activity of dorsal PFC (dPFC) neurons at a given time bin between 0 and 800 ms after cue onset was correlated with the activity of ACC neurons in subsequent time bins, suggesting that dPFC neurons play a role in driving ACC neurons. Third, the ACC–PFC correlation was statistically significant only when spike trains were smoothed using Gaussian kernels with widths of 50–200 ms. Since this time-scale is far larger than that of monosynaptic communication, it suggests that attention control involves coordination of large-scale brain networks.

To quantify interneuronal interactions, Oemisch and colleagues (2015) calculated correlations between the instantaneous activities of neuron pairs. This measure quantifies, for a given instant from cue onset, the correlation of spike rates for a given neuron pair across trials. However, their behavioral paradigm might create spurious intertrial correlations: given that the response of a considerable proportion of neurons in the frontal cortex depends on the visual properties of behaviorally relevant stimuli (Zaksas and Pasternak, 2006; Bichot et al., 2015), the response of these neurons will change when attention is switched between the left and right visual hemifield

Received Nov. 23, 2015; revised Dec. 25, 2015; accepted Dec. 30, 2015.

This work was supported by the Iranian Cognitive Sciences and Technologies Council. We thank Stefan Treue and Mohammad Reza Daliri for their kind comments on an earlier version of the article.

The authors declare no competing financial interests.

*M.E. and C.X. contributed equally to this work.

Correspondence should be addressed to Moein Esghaei, Cognitive Neuroscience Laboratory, German Primate Center (DPZ), Kellnerweg 4, Goettingen 37077, Germany. E-mail: esghaei@ipm.ir.

DOI:10.1523/JNEUROSCI.4211-15.2016

Copyright © 2016 the authors 0270-6474/16/361791-03\$15.00/0

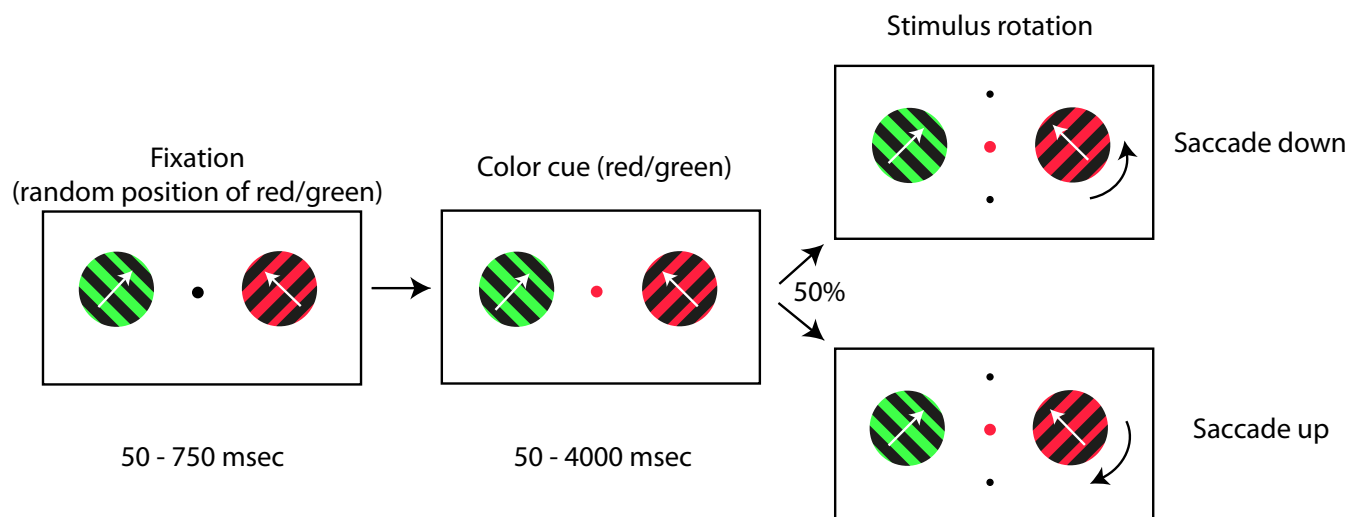


Figure 1. Behavioral task used by Oemisch et al. (2015). Monkeys had to maintain their gaze on a central point while a $+45^\circ$ and a -45° drifting grating were presented to the left and right visual hemifield, respectively. In each trial, one of the gratings was randomly selected to be green and the other, red. At a random time between 50 and 750 ms, the fixation point changed color to red or green, cueing the monkeys to covertly shift their attention to the matching (target) grating. After a random period of 50–4000 ms, one of the gratings underwent a clockwise or counterclockwise change in direction of motion. Monkeys were trained to make an upward or downward saccade if the target grating underwent a clockwise or counterclockwise change in direction, respectively. The white arrows represent drifting direction and were not shown in the actual experiment (details originally described by Kaping et al., 2011).

(to gratings drifting at $+45^\circ$ and -45° , respectively). This alone can explain the trial-by-trial correlation induced after cue onset when the visual information about the attended stimulus has reached the neurons: assuming a pair of neurons that are both selective to the contralateral grating (which plays a major role in the main finding of the paper; Oemisch et al., 2015, their Fig. 5A), both will increase their firing rate in trials on which the monkeys attend contralaterally, and decrease their firing rate when the animals attend ipsilaterally, thus creating an intertrial correlation. The influence of such an effect can be avoided by only analyzing trials in which the monkeys attend to the same drifting direction. Similarly, when the authors compare the interneuronal correlation between contralateral and ipsilateral attention shifts (Oemisch et al., 2015, their Fig. 5Biii), any difference can be accounted for by the difference in signal-to-noise ratio of the neural response to the preferred versus nonpreferred stimulus. This could be controlled by computing the correlations for each neuron, using a subset of trials in which the neuron does not show a significant change in spike rate between contralateral/ipsilateral attention shifts. Alternatively, future experiments could present two identical stimuli in the two visual hemifields to induce the same neural responses between contralateral/ipsilateral attention shifts. Whether the interneuronal correlation will remain statistically significant after removing the component induced by differential visual stimulations is an important concern,

however, given the small correlation values (average Pearson correlation coefficient < 0.02) (Oemisch et al., 2015, their Fig. 2B) compared with similar studies in sensory areas (Mitchell et al., 2009).

To investigate the role of attention-evoked neural correlation between ACC and dPFC neurons, Oemisch et al. (2015) quantified the directionality of the interneuronal interaction using a directionality index, proposed earlier (Paz et al., 2007). This index measures, for a given pair of neurons (A and B), the difference between the mean trial-by-trial correlation of instantaneous neural activities across time-pairs with the activity of neuron A preceding neuron B and those time-pairs with the activity of neuron B preceding neuron A. This gives a measure of whether either of the neurons in a given pair activated the other neuron. However, this method suffers from dependence on the time scale it is applied to. As the authors point out, when changing the analysis period to 300–600 ms after cue onset (rather than 0–800 ms), the directionality effect disappears unless the spike trains are smoothed by convolving an extra-large Gaussian kernel with the width of 400 ms rather than 50–200 ms. Therefore, to study the directionality of interneuronal interactions in a given interval, all constituent time windows should be investigated.

Oemisch et al. (2015) found that interneuronal correlations only emerged when the Gaussian kernel width used for smoothing spike trains was larger than 50 ms (Oemisch et al., 2015, their Figs. 1C,

2A, 3C), with an understandable exception of 400 ms kernel width because it expanded beyond the analysis window (Oemisch et al., 2015, their Figs. 1D, 2E, 3C). The authors interpret this as a characteristically slow time scale for coordination across areas. Yet the underlying mechanism could easily be further explored. The most straightforward interpretation of the effect of kernel width on firing correlation is that attention induced a certain pattern of neuronal events, which by themselves have stochastic time intervals that distribute ~ 50 ms or longer; therefore, a consistent correlation across areas becomes evident only when a time window larger than 50 ms is used to overcome the stochasticity. Meanwhile, another study on the same dataset (Womelsdorf et al., 2014) demonstrated that there is an increase in the proportion of neuronal bursts (clusters of spikes intercepted with variable quiescent periods) after the attention cue onset. They further found that these bursts are synchronized with the narrow beta band (12–20 Hz) of local field potentials (LFP), which, probably not coincidentally, corresponds to a period of > 50 ms. Therefore, we speculate that the characteristic slow time scale reflects the stochasticity between the bursts, which occurs within a period of beta-band rhythm. This hypothesis can be directly tested by looking into potential correlations between attentional effects on burstiness and interneuronal correlation. The authors made an attempt along this line when they reported no across-cell-pair correlation between burst

proportions and correlation strengths. However, this lack of correlation is probably not surprising, given that different cell types in prefrontal cortex have distinct burst properties to start with, independent of their correlation induced by attention (shown on the same dataset by Ardid et al., 2015). Therefore, a more relevant test would address the link between the two effects within cell pairs, i.e., whether for a given neuron pair, the attentional modulation of neural correlation is stronger when calculated for bursty trials compared with less bursty trials. If there is no link between interneuronal correlation and burstiness, interneuronal correlation could be attributed an independent role in controlling attention.

In summary, Oemisch et al. (2015) provided evidence suggesting the involvement of interneuronal interaction between areas ACC and dPFC when monkeys switched their attention to different entities. They also suggested that dPFC neurons control the activity of ACC neurons when deploying attention. However, there are at least three more steps to accomplish before reaching a firm conclusion about the role of interneuronal correlations in controlling attention: first, to disentangle the interneuronal correlation from differential neuronal responses evoked by stimulus variants; second, to in-

vestigate all constituent time windows of postcue period for any potential role of dPFC in governing ACC for attention control; and third, to clarify the relationship between attentional modulations of correlated firing and burstiness–LFP synchronization.

References

- Ardid S, Vinck M, Kaping D, Marquez S, Everling S, Womelsdorf T (2015) Mapping of functionally characterized cell classes onto canonical circuit operations in primate prefrontal cortex. *J Neurosci* 35:2975–2991. [CrossRef Medline](#)
- Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. *Science* 344:424–427. [CrossRef Medline](#)
- Bichot NP, Heard MT, DeGennaro EM, Desimone R (2015) A source for feature-based attention in the prefrontal cortex. *Neuron* 88:832–844. [CrossRef Medline](#)
- Gregoriou GG, Gotts SJ, Zhou H, Desimone R (2009) High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* 324:1207–1210. [CrossRef Medline](#)
- Kaping D, Vinck M, Hutchison RM, Everling S, Womelsdorf T (2011) Specific contributions of ventromedial, anterior cingulate, and lateral prefrontal cortex for attentional selection and stimulus valuation. *PLoS Biol* 9:e1001224. [CrossRef Medline](#)
- Maunsell JH, Treue S (2006) Feature-based attention in visual cortex. *Trends Neurosci* 29:317–322. [CrossRef Medline](#)
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202. [CrossRef Medline](#)
- Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63:879–888. [CrossRef Medline](#)
- Moore T, Armstrong KM (2003) Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421:370–373. [CrossRef Medline](#)
- Oemisch M, Westendorff S, Everling S, Womelsdorf T (2015) Interareal spike-train correlations of anterior cingulate and dorsal prefrontal cortex during attention shifts. *J Neurosci* 35:13076–13089. [CrossRef Medline](#)
- Paz R, Bauer EP, Paré D (2007) Learning-related facilitation of rhinal interactions by medial prefrontal inputs. *J Neurosci* 27:6542–6551. [CrossRef Medline](#)
- Petersen SE, Posner MI (2012) The attention system of the human brain: 20 years after. *Annu Rev Neurosci* 35:73–89. [CrossRef Medline](#)
- Treue S (2001) Neural correlates of attention in primate visual cortex. *Trends Neurosci* 24:295–300. [CrossRef Medline](#)
- Womelsdorf T, Ardid S, Everling S, Valiante TA (2014) Burst firing synchronizes prefrontal and anterior cingulate cortex during attentional control. *Curr Biol* 24:2613–2621. [CrossRef Medline](#)
- Zaksas D, Pasternak T (2006) Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J Neurosci* 26:11726–11742. [CrossRef Medline](#)

Chapter 7

Summary and discussion

This thesis endeavored to address scientific questions of the neuronal mechanism of visual perception at various stages, from the sensory end to the motor end. From the sensory end, I looked into how multiple dimension of visual information from a stimulus (such as the 3D location, and the moving direction of a moving pattern) are represented in neuronal activity in macaque visual cortical area MST (chapter 2). I then investigated which aspects of the neuronal activity in macaque MST are further modulated by directing attention to a location or feature (chapter 3). Apart from its effects on neuronal activities in visual cortices, on the motor end, directing attention to a spatial location is also correlated with the direction of involuntary eye movements in human subjects (chapter 4). And finally, I presented in chapter 5 a direct comparison between the voluntary eye movement patterns of humans and monkeys while free-viewing images, to assess to which extent attentional selection mechanism is shared across the two species.

It is an intriguing question how and why neurons encode multiple features simultaneously. Previous studies in macaque visual cortical area MST have reported neurons with selectivity to the direction of a moving stimulus (e.g. 'to which direction is the car going?'), as well as the distance between the stimulus relative to the gaze horopter (i.e. 'how far away is the car?'). In chapter 2, I discussed the possibility that information about both motion direction and depth are encoded in area MST because the interaction of the two is important for a certain cognitive function, such as self-motion perception. To our surprise, our data showed that the representation of the motion direction and depth are largely independent. This suggests that area MST encodes the information of motion direction and depth in a separate manner, while

the integration of the two and the consequential self-motion perception may be found in other visual areas, presumably higher in the hierarchy than MST.

The sensory representation of visual stimulus is further modulated by attention, which selectively enhance a specific location or a stimulus feature according to the behavioral context. The results in chapter 2 implies that stimulus features and stimulus location are separately represented in MST, as if location is simply one of the many features of the visual object (e.g. shape, color, movement direction etc.). To test whether space is just another feature to the brain, I showed in chapter 3 a study that investigated whether directing attention to a location induce similar effects as directing attention to a feature. As a result, spatial attention induced a reduction in neuronal burstiness, while feature-based attention didn't. This suggests that directing attention to spatial locations involves a unique neuronal mechanism compared with directing attention to features, and that the visual system in the brain treats information about space and feature differently.

It is beyond the scope of this thesis to find out why spatial attention is different from feature-based attention. But on the behavioral level, spatial attention is unique in its tight correlation with the eye movement system. A popular hypothesis equated covert spatial attention (directing attention to a location without voluntarily moving the eyes) with the planning of an eye movement to that location. In chapter 4, I showed strong evidence in human subjects, that although the subjects are prohibited from moving their eyes, the directions of the minute involuntary eye movements (microsaccades) are biased toward the attended locations. This effect can be isolated from confounding effects from the cue contingency and motor intentions of the subjects. Such a reliable correlation between microsaccades and attention suggests, that the attentional state of a human subject can be inferred simply from his/her involuntary eye movement pattern.

As the knowledge about the human visual perception (including findings in this thesis) mainly come from a combination of studies in human behavioral studies and monkey neurophysiological studies, it is a critical issue how much the two species share in attention selection mechanisms. In chapter 4, I showed a study I contributed to, which compared the two species in terms of the patterns of eye movement and fixation during freely viewing of images. Despite the vastly different behavioral repertoires of the two species, our analysis demonstrated a surprising level of similarity between the scan paths of humans and monkeys when they look at the same set of pictures. Apart from differences in response to presumably high-level features (such as 'interestingness'), the features that attract monkeys' attention also attract that of humans. This finding confirms that the principles of visual information processing we discovered so far by electrophysiology in monkeys are largely applicable to humans as well.

Taken together, the findings of this thesis contributed to various aspects of the current literature on primate visual perception, especially on the topics of how visual information is represented and modulated according to behavioral needs. In a broader context, the findings we obtained with our specific experiments (e.g., in cortical area MST, with motion dot stimuli, with experiment designed to investigate attention) provided a unique perspective to the general principles of visual information processing in primates. Other independent studies (in different cortical areas and different experiment designs) are necessary to confirm and further contribute to our understanding of the neuronal mechanism of vision.

Curriculum Vitae

Cheng Xue

AREA OF INTEREST: Systems Neuroscience

AFFILIATION:

Department of Neuroscience
University of Pittsburgh

EMAIL ADDRESS: chx31@gwdg.de

GENDER: Male

INSTITUTE ADDRESS:

Center for the Neural Basis of Cognition,
115 Mellon Institute,
4400 Fifth Avenue,
Pittsburgh, PA, 15213
United States

PHONE: +86 18261578919

NATIONALITY: P. R. China

BORN: Jan. 6, 1987

EDUCATION: Sep. 2012- Oct. 2016: Ph.D. student in Systems Neuroscience, GGNB graduate school, Georg-August-University Göttingen, Germany.

Sep. 2011- Sep. 2012: Research Assistant (Systems Neuroscience), Cognitive Neuroscience Laboratory, German Primate Center;

Sep. 2009- Jul. 2011: Master of Science (Biophysics), Department of Physics, Nanjing University, China;

Sep. 2005- Jul. 2009: Bachelor of Science (Physics), Department of Intensive Instruction, Nanjing University, China;

PUBLICATIONS AND CONFERENCE CONTRIBUTIONS:

JOURNAL ARTICLES:

- **Xue C**, Liu F (2014) *Structured Synaptic Inhibition Has a Critical Role in Multiple-Choice Motion-Discrimination Tasks* **J. Neurosci.** 34(40): 13444-13457
- Esghaei M, **Xue C** (2016) *Does correlated firing underlie attention deployment in frontal cortex?* **J. Neurosci.** 36(6): 1791-1793
- **Xue C**, Krishna BS, Kaping D, Baloni-Ray S, Treue S (2016) *Spatial, but Not Feature-Based Attention, Reduces Burstiness in Area MST of the Macaque Monkey* **Cerebral Cortex** 27 (1): 83-91, doi: 10.1093/cercor/bhw326.
- Wilming N, Kietzmann T, Jutras M, **Xue C**, Treue S, Buffalo E, König P *Differential contribution of low and high-level image content to eye movements in monkeys and humans.* **Cerebral Cortex**, doi: 10.1093/cercor/bhw399
- **Xue C**, Calapai A, Dannhaeuser K, Krumbiegel J, Treue S *Sustained spatial attention alone explains the bias in microsaccade direction, under preparation*
- Calapai A, **Xue C**, Treue S *Independent visual-motion and binocular disparity in Macaque Monkeys area MST, under preparation*

THESIS:

- **Xue C** (2016) *Neuronal representation and attentional modulation of space and feature information in primate vision modelling Neural Circuit for Spatial Working Memory and Perceptual Decision Making*, Ph. D. thesis ('*summa cum laude*'), Georg-August-University Göttingen, Germany
- **Xue C** (2011) *Modelling Neural Circuit for Spatial Working Memory and Perceptual Decision Making*, M. Sc. thesis, Nanjing University, Nanjing.

- **Xue C** (2009) *The Neural Dynamics in a Decision-Making Circuit*, B. Sc. thesis, Nanjing University, Nanjing

ACADEMIC POSTERS:

- **Xue C**, Calapai A, Krumbiegel J, Treue S (2016) *Spatial attention, by itself, explains the direction bias in human microsaccad-direction*. Society for Neuroscience Annual Meeting 2016, San Diego
- **Xue C**, Krishna BS, Baloni-Ray S, Kaping D, Treue S (2015) *Spatial attention, but not feature-based attention, is correlated with reduction in bursty firing in area MST*. Society for Neuroscience Annual Meeting 2015, Chicago
- **Xue C**, Krishna BS, Baloni-Ray S, Kaping D, Treue S (2015) *Spatial attention, but not feature-based attention, is correlated with reduction in bursty firing in area MST*. Primate neurobiology meeting, Göttingen
- **Xue C**, Calapai A, Dannhäuser C, Treue S (2014) *The correlation between fixational eye movement and visual spatial attention*, European Summer School for Eye Movement, Freiburg
- **Xue C**, Liu F (2010) *A Cortical Loop Model Underlying Fine Discrimination of Motion Direction* at the Cold Spring Harbor Francis Crick Neuroscience Symposium, Suzhou

TRAINING EXPERIENCE:

GENERAL SCIENTIFIC METHODS:

- May. 2014: Course *Visualization of scientific data* (GGNB), Göttingen
- Oct. 2013: Course *Basic statistics for graduate students in the life sciences* (GGNB), Göttingen

SYSTEMS NEUROSCIENCE:

- Apr.-Jul. 2015: Course *Theoretical and computational neuroscience* (GGNB), Göttingen
- Sep. 2014: *European Summer School for Eye Movement* (organized by Prof. Dr. Christoph Klein, Prof. Dr. Ulrich Ettinger), Freiburg
- May. 2013 – Jul. 2013: Course *Visual Psychophysics – From Theory to Experiment* (given by Prof. Dr. Stefan Treue and Dr. Clíodhna Quigley), Göttingen
- Sep. 2012: *10th Summer Course on Computational Neuroscience*, (organized by Max Planck Institute for Dynamics and Self-Organization), Göttingen
- Oct. 2011 – Feb. 2012: Course *From Vision to Action* (given by Prof. Dr. Stefan Treue, Prof. Dr. Alexander Gail, and Prof. Dr. Hansjörg Scherberger), Göttingen
- May 2011: *Tutorial on patch clamp technique and application* (given by Prof. Erwin Neher), in Nanjing Univ., Nanjing
- July 2010: *CSHA Summer School on Simulation and Modelling in Neuroscience*, Suzhou (organized by Prof. Xiao-Jing Wang and Prof. Si Wu)
- May 2010: *SJTU Short-term Training on Computational Neuroscience*, Shanghai (organized by Prof. Xiao-Jing Wang and Prof. Si Wu)
- Aug 2009: *IBRO-APRC Associated School* (organized by Prof. Jian-Jun Wang) in Nanjing Univ., Nanjing

TECHNIQUES AND ETHICS OF ANIMAL EXPERIMENT:

- Mar. 2016: Course *Primate neurobiology methods: behavior, experiments, analysis, and ethics* (GTPN), Tübingen

- Jun. 2015: Course *Laboratory Animal Science Course on Primates*, Göttingen
- Mar. 2013: Course *Introductory workshop as part of Graduate Training in Primate Neuroscience (GTPN)*, Tübingen
- Apr. 2012: Course *Versuchstierkundlicher Einführungskurs (Introduction to the field of animal experiments)*, Max-Planck Institute of Biophysical Chemistry, Göttingen
- Nov. 2011: *Course on General Primate Biology* (organized by EUPRIM-Net), Göttingen

PROFESSIONAL SKILLS:

- Mar. 2016: Workshop *Leading without a leadership position* (GGNB), Göttingen
- May 2015: Workshop *Social Networking Online for Career Purposes* (GGNB), Göttingen
- Apr. 2015: Workshop *Project management* (GGNB), Göttingen
- Feb. 2015: Workshop *Global Leadership in the Knowledge Society* (GGNB), Göttingen
- Jan. 2015: Workshop *Effective scientific communication: journal papers, seminar or conference presentations, and posters* (offered by GGNB), Göttingen
- Dec. 2014: Seminar *Career planning: what to do with a Ph.D.?* (GGNB), Göttingen
- May. 2014: *Seminar on Good Scientific Practice* (GGNB), Göttingen
- Mar. 2014 – June 2014: Workshop *Writing Scientific Publications* (GGNB), Göttingen
- Oct. 2013: Workshop *Team work & leadership competencies in academia and beyond* (GGNB), Göttingen
- Oct. 2013: Workshop *Networking for academics* (GGNB), Göttingen
- Jun. 2013: Seminar *Good Scientific Practice* (organized by Prof. Julia Fischer, German Primate Center), Göttingen

TEACHING EXPERIENCE:

- Jan. 2013 – Feb. 2013: supervisor for internship (Kristin Dannhäuser), Göttingen
- Apr. 2013 – Oct. 2013: supervisor for Master thesis (Kristin Dannhäuser), final grade 1.0 (top), Göttingen
- Mar. 2014 – May. 2014: supervisor for lab rotation (Michael Siebrecht), Göttingen
- Apr. 2015 – May. 2015: supervisor for internship (Julius Krumbiegel), Göttingen
- May. 2015 – Aug. 2015: supervisor for Bachelor thesis (Julius Krumbiegel), final grade 1.3 (good), Göttingen

Many duties are entailed in supervising internship / lab rotation, Bachelor thesis, and Master thesis. Apart from training students on data collection / analysis skills, an equally important part is to work with the students to improve the quality of the final written report / thesis, upon which their final grading is based. All the three projects I supervised received top or good grading.

- Apr. 2016: tutor of seminar on attention for master students of the Neuroscience Program, Göttingen
- Apr. 2015: tutor of seminar on attention for master students of the Neuroscience Program, Göttingen
- Jan. 2013: tutor of seminar of the lecture *Vision to Action*, Göttingen
- May. 2010 – Aug. 2010: Teaching assistant for undergraduate course College Physics Experiment, Nanjing University, Nanjing

CONFERENCES ATTENDED:

- Nov. 2016: Annual Meeting of the Society for Neuroscience, San Diego.
- Mar. 2016: 9th Primate neurobiology meeting, Tübingen
- Oct. 2015: Annual Meeting of the Society for Neuroscience, Chicago.
- Mar. 2015: 11th Meeting of the German Neuroscience Society, Göttingen.
- Mar. 2015: 8th Primate neurobiology meeting, Göttingen
- Mar. 2014: 7th Primate neurobiology meeting, Tübingen
- Dec. 2013: GGNB Science Day, Göttingen.
- Mar. 2013: 6th Primate Neurobiology Meeting, Göttingen.
- Mar. 2013: 10th Meeting of the German Neuroscience Society, Göttingen.
- Nov. 2012: Conference Orienting of Attention, Neural Implementation, Underlying Mechanisms and Clinical Implications, (organized by Division of Neuropsychology, University of Tübingen), Tübingen.
- Mar. 2012: 5th Primate Neurobiology Meeting, Tübingen.

OTHER ACTIVITIES:

- Sep. 2005-Apr. 2007: Editor for annually campus-based literary periodical *Table Talk*, Nanjing University
- Sep. 2005-Apr. 2007: Science Editor for annually campus-based journal *Forum of basic sciences*, Nanjing University
- Apr. 2007-Dec. 2008: Secretary and finance manager of Matial Art Association, Nanjing University
- Sep. 2007-Dec. 2010: Volunteer worker at Boai Asylum for mentally challenged youth, Nanjing. (Awarded 2010 ‘volunteer of the year’ by Nanjing Charity Foundation)
- Nov. 2012: Participated in organizing exhibition for the first *Nacht des Wissens* (night of science), Göttingen
- Jan. 2016: Participated in organizing exhibition for the third *Nacht des Wissens* (night of science), Göttingen

PERSONAL SKILLS:

Hardware knowledge: eye tracking systems, electrophysiology setup

Software Skills: Neuroscience related softwares: Osrix (MRI scan); MWorks (experiment control and data collection); Plexon (neuronal data collection and offline sorting)

Programming Language: C, Matlab;

Natural Language: Chinese (native), English (fluent), German (level B2)