



HAL
open science

Behavioral and functional neuroimaging investigation of motion integration in early visual cortical hierarchy

Kim Beneyton

► **To cite this version:**

Kim Beneyton. Behavioral and functional neuroimaging investigation of motion integration in early visual cortical hierarchy. Neuroscience. Université Claude Bernard - Lyon I, 2023. English. NNT : 2023LYO10028 . tel-04555050

HAL Id: tel-04555050

<https://theses.hal.science/tel-04555050>

Submitted on 22 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESE de DOCTORAT DE L'UNIVERSITE CLAUDE BERNARD LYON 1

**Ecole Doctorale N° 476
Neurosciences et Cognition**

Discipline : Neurosciences

Soutenue publiquement le 28/02/2023, par :
Kim Beneyton

Behavioral and functional neuroimaging investigation of motion integration in early visual cortical hierarchy

Devant le jury composé de :

Derrington, Edmund	Prof. des Universités, Université Lyon 1	Président
Goebel, Rainer Montagnini, Anna	Prof., Université de Maastricht (Pays-Bas) Dir. de Recherche, CNRS Marseille	Rapporteur Rapporteuse
Bonnefond, Mathilde Collins, Thérèse Rademaker, Rosanne	Chargée de Recherche-HDR, INSERM Lyon Prof. des Universités, Université Paris Cité Prof., ESI for Neuroscience, Frankfurt (Allemagne)	Examinatrice Examinatrice Examinatrice
Knoblauch, Kenneth Kennedy, Henry	Dir. de Recherche Emérite, INSERM Lyon Dir. de Recherche Emérite, INSERM Lyon	Directeur de thèse Co-directeur de thèse

ACKNOWLEDGEMENTS

After four years and a few months,
And for what it is worth,
It is time to deliver my words,
And to thank my world.
I am surrounded by authentic persons,
Who fed me with inspiration.
No one resembles the other.
Neither in shape nor in color.

I am grateful to my team.
And I missed it when I could not see them.
Any source of happiness becomes more vivid
After having been deprived.
Coach Henry was the first I met,
Every step of the way, every stop on my track,
He would find the words, give a pat on my back,
And the athlete goes for another round.
On my way I'd see Ken at the next turning point,
Handing me a glass of water, a towel and biscuits,
Always there to remind me of the hidden statistics,
A poet of the numbers and mathematics.

I cannot mention all the staff
I've met along this path,
But they were my everyday,
Going through the same journey.
How many times helping
Without knowing?
Add a few jokes in the background,
Such a fertile ground.

Suddenly, I would feel the urge to go out,
Ride my bike, Breath out.
Thereby meeting my beloved ones,
The never-ending loop of strong friends.

I cherish my family more than anyone else
In my blood, the red thread, what else?
Mom, Dad, & my Fantastic 3, I love you.
And they are so many more... Thank you.

Lastly, I would like to thank Hip-Hop for making me feel alive
and Neuroscience for allowing me to think about life.

ABSTRACT

The work of my thesis is best understood in the context of the hierarchical organization of the visual system, the functional consequences of hierarchy and its role in perceptual decisions. Accordingly, the purpose of this introduction is to review the structural and functional specificities of feedforward (FF) and feedback (FB) streams in the cortex and how they relate to hierarchical processing. Unlike FF function that has been extensively studied and associated to the building of increasingly complex representations of the outside world, much less is known about the functional and perceptual roles of FB signals in cortico-cortical networks.

Models of motion integration are ideal candidates for reviewing functional hierarchy in early visual cortex. Motion integration has been extensively studied and constitutes an influential model originally represented as a two-stage feedforward process where FB signals played no apparent role in motion perception. Within this model, low-level visual areas are proposed to linearly detect and encode one-dimensional (or “component”) motion (e.g., a single grating moving orthogonally to its orientation), while the higher-level motion detectors compute non-linear integration of more complex (or “pattern”) motion signals (e.g., two superimposed moving gratings integrated as one moving plaid). According to this view, low-level visual areas (primary and secondary visual areas, V1 and V2) and higher-level ones (e.g., MT/V5, middle temporal area) represent motion information differently. However, more recent data from single-cell responses and cortical imaging in macaque indicate that pattern motion signals can be recorded as early as area V1. My thesis addresses whether such signals can be detected in early visual areas in human cortex and considers whether they might arise from feedback processes.

Given the large hierarchical distance between them, V1 and MT/V5 have the particularity to be relatively strongly anatomically interconnected. This suggests that V1-MT/V5 coupling might play an important role in motion perception. This idea is supported by a number of studies indicating that MT-to-V1 FB connectivity might play a key role in generating mental representations of perceived motion based on incomplete or illusory sensory input. Interestingly, the relation between these two areas can be investigated through a well-known paradigm of bistable motion which consists of superimposed moving gratings of different orientations that can generate multiple perceptual states over time. These perceptual states appear as spontaneous shifts between the perception of a coherent plaid pattern moving in a single direction and that of a pair of component gratings, transparently sliding over each other in two different directions. Several neurocomputational models have been proposed to explain the neural mechanisms underlying bistable motion perception. In general, they propose that the perception of ambiguous motion signals can be explained by a competition between neural populations that support distinct representations of motion. For example, one hypothesis states that these neural populations are composed of both excitatory and inhibitory units that are involved in both competitive and adaptative mechanisms over time.

In my thesis, after designing an efficient bistable moving plaid illusion, we used an event-related fMRI paradigm to study bistable perception dynamics. In preliminary measures, we were able to identify direction-selective voxels in early visual areas V1, V2 and the hMT+ complex (human motion complex, hMT+, equivalent to MT/V5 in non-human primate and adjacent motion-selective areas). We went on to engineer conditions that would trigger spontaneous perceptual switches while maintaining a constant visual stimulation. The two competitive perceptual states are proposed to consist of two equiprobable and mutually exclusive representations of motion information. While subjects fixated a central cross, to

minimize eye movements, they were asked to report their perceptual state. Analyses of the hemodynamic activity in these areas demonstrated that the direction-selective voxels preferentially responded to the perceptual state in accordance with the matching direction. In subsequent analyses, we argue that the results represent neural activity related to the perceptual switching, and are not reflecting eye movement phenomena.

TABLE OF CONTENTS

CHAPTER 1: Introduction	8 - 35
1. Anatomy and function of the visual cortical hierarchy	
1.1 Historical overview	
1.2 Anatomy of Hierarchy	
1.3 Towards the definition of a functional hierarchy	
2. Investigating FB function through perceptual integration	
3. A well-known model of visual hierarchy: Motion integration	
3.1 Motion signals	
3.2 Moving plaids and The Aperture Problem	
3.3 Perceptual competition: from general to specific rules	
3.4 Neurocomputational models of bistable perception	
4. Summary	
CHAPTER 2: The Dynamics of Bistable Perception	36 - 50
1. Psychophysics: How to design a bistable plaid?	
1.1 General properties of bistable plaids	
1.2 Motivations	
1.3 Methodology	
1.4 Introduction to binomial Generalized Linear Model (glm)	
2. Group-level statistics for the additive model	
3. Conclusion	
CHAPTER 3: Neural Correlates of Perceptual Decisions	52 - 82
1. Introduction to functional magnetic neuroimaging	
1.1 Biology	
1.2 Physics	
2. Modeling BOLD signal with the General Linear Model	
2.1 Introduction to GLM statistics	
2.2 Multilevel GLM analyses	
3. Material & Method	
4. Analyses pipeline	
4.1 Behavioral data	
4.2 fMRI preprocessing	
4.3 fMRI analysis: GLM	

4.4 Group-level statistics: Linear Mixed Effect Model

5. Results

5.1 Behavioral results

5.2 Estimate of Pattern/Component selectivity

6. Conclusion

CHAPTER 4: Analysis of a potential role of Eye Movements in Bistable Motion Perception

84 - 105

1. Effect of eye movements in bistable perception

1.2 From sensing to perceiving

1.2 Eye control and perception in visual cortex

2. Material & Method

3. Analyses pipeline

3.1 Analysis 1: Predicting perceptual state from previous eye movements

3.2 Analysis 2: Comparison Localizer/Bistable perception

4. Results

5. Conclusion

CONCLUSION & PERSPECTIVES.

106 - 113

BIBLIOGRAPHY

114 - 138

SUPPLEMENTARY

140 - 175

CHAPTER I
Introduction

1. Anatomy and function of the visual cortical hierarchy

1.1 Historical overview

The notion of hierarchical processing in the visual cortex was introduced by Hubel and Wiesel to account for the variation in receptive field (RF) organization in early visual processing. They conducted a series of experiments involving the presentation of oriented and non-oriented stimuli while recording receptive field properties in the thalamus and visual cortex in both cats (1962; 1965) and macaques (1968). They proposed a hierarchical organization which amounted to a serial model of processing whereby the so-called simple cells in layer 4 of cortical area V1 of the cat exhibited oriented RFs that were constructed from the summation of the RFs of non-oriented cells in the Lateral Geniculate Nucleus (LGN) and where neurons in area V1 (displaying simple RFs) converged to the so-called complex cells of area V2 (exhibiting a more “complex” RF organization). These authors envisioned that this process achieved a complexification of stimulus selectivity. In their model, cells at early stages of vision were integrated hierarchically to generate selectivity to more complex features at subsequent stages of processing. These discoveries were fundamental to understanding functional specificities of neurons in early visual cortex.

Later, Rockland and Pandya (1979; 1981) described a hierarchical model of inter-areal connectivity with ascending stream (feedforward, FF) going from lower to higher-order areas, and descending connections (feedback, FB) going from higher- to lower-order areas. They proposed that FF were defined by their projection to layer 4 (by analogy to the projection of LGN to layer 4 of V1), and FB by their avoidance of layer 4. In 1991, Felleman and Van Essen exploited this categorization of FF and FB and provided a more general model of visual hierarchy, where the increase in the number of connections led to numerous parallel pathways, hence characterizing a distributed hierarchy. Their analysis was based on a comprehensive review of the literature (up to 1990) of antero- and retro-grade tract tracing studies, investigating the laminar patterns of inter-areal connectivity in macaque cortex. More recently, quantitative measures of the hierarchical connectivity reveal that FB pathways are more numerous and cross more hierarchical levels than do FF pathways (Markov et al., 2014b).

Although Felleman & Van Essen’s hierarchy provided important insight into the cortical anatomy it was indeterminate as it yielded over 150,000 possible hierarchical orderings (Hilgetag et al., 1996). This problem was overcome by quantifying the laminar, inter-areal connectivity and establishing a metric of hierarchical distance based on retrograde tract-tracing (Barone et al., 2000; Markov et al., 2014b; Vezoli et al., 2021), thereby defining a determinate model of hierarchy. Further, the investigation of the cortical anatomy of visual hierarchy led to the identification of two distinct FF and FB pathways located in the infragranular and supragranular compartments (Markov et al. 2014b; Vezoli et al., 2021). These multiple FF and FB patterns of connectivity led to the concept of a Dual Counter Stream architecture (Fig. 2), defining specific FF and FB relations in the two compartments based on an index of hierarchical distance. This model is increasingly influential for understanding diverse aspects of hierarchical processing in the cortex of rodents (Barzegaran & Plomp, 2022) and primates (Markov et al., 2013b; Vezoli et al., 2021; Maric & Domijan, 2022).

1.2 Anatomy of hierarchy

Simple and Complex neurons

Visual input originates from the retina, where photoreceptors transduce light into chemo-electrical activity. At the retinal output, ganglion cells transmit the signals to the thalamic nucleus, the Lateral Geniculate Nucleus (LGN), which, in turn, relays signals to the first cortical stage of vision, the primary visual area (area V1). Area V1 occupies a unique position in the visual cortical hierarchy as it receives no FF cortical input and has no FB projections to other cortical areas. Interestingly, transcriptomic cytoarchitecture across human cortex shows that area V1 exhibits dramatically different features compared to other cortical areas, a specificity that could have important impacts on the FB pathways targeting area V1 (Jorstad et al., 2022 bioRxiv). Furthermore, area V1 is regarded as a mandatory element of passage where removal of area V1 in human subjects leads to a loss of conscious vision and a residual visual capacity known as blindsight, which is thought to involve subcortical direct visual input to MT/V5 (see review Bourne & Morrone, 2017).

In 1962, David Hubel and Torsten Wiesel were awarded the Nobel prize in part for their work describing the functional organization and basic physiology of neurons in V1 (cat: Hubel & Wiesel, 1962; 1977; monkey: 1968; 1976). In terms of functional architecture, area V1 in cats and macaques is composed of columnar units at two different scales, corresponding to orientation and ocular dominance columns. Area V1 responds to many different stimulus properties such as color, spatial frequency, motion direction etc. In this section, we focus on the elementary features that present a clear anatomical and functional segregation in area V1.

Hubel and Wiesel distinguished three main types of neurons on the bases of their RF properties and activity profiles. *Simple cells*, like retinal ganglion (RGC) and LGN cells, are characterized by separate zones of ON and OFF responses to light. The integrated response of a simple cell corresponds to the spatial distribution of these ON and OFF regions (Fig. 1). Hubel and Wiesel hypothesized that simple cells sum input signals from multiple LGN cells having adjacent and aligned RFs, thereby conferring upon each cell a particular orientation preference. Hence, unlike the RGC and LGN cells that have a concentric center-surround organization, the simple cells respond best to oriented lines or edges, which in mammals is a hall mark of the early visual cortex (but see Douglas et al., 1995; 1989; Ferster et al., 1996).

Simple cell sums LGN inputs

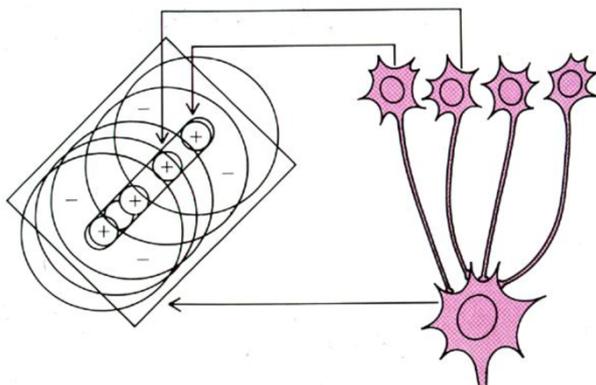


Figure 1: Illustration by David Heeger, 2006

Hubel and Wiesel went on to propose that *complex cells* non-linearly integrate input signals from multiple, similarly oriented simple cells that present distinct polarities. Therefore, complex cells do not display separate ON and OFF zones, but instead present a spatially homogeneous RF. As a consequence, complex cells reflect the same preferences as simple cells (orientation selectivity), except that a complex cell's response does not depend on the spatial location of a stimulus within its RF (phase insensitivity). In addition, complex cells tend to be more direction-selective compared to simple cells.

Hypercomplex cells share similar properties with complex cells except that they contain inhibitory flanks at the spatial extremities of their RFs, so that while their response to a bar stimulus initially increases with length, beyond a certain length it decreases due to the presence of inhibitory flanks. This last property referred to as end-stopping results in the cell maximally responding to a short-oriented line segment or a more complex two-dimensional feature, such as a corner. Note that there is a category of simple-like cells that also show this end-stopping property (Orban & Kennedy, 1981).

Based on their relative properties, these three cell types enable processing of different and successively more complex features of the visual environment. This is the basis of many hierarchical model of how compound forms can be represented by combinations of cells responding to simpler features (Barlow et al., 1972).

Receptive field or the expression of FF processing

Ascending visual pathway relaying retinal input to area V1 preserve retinotopic relations whereby neighboring positions in the visual field are encoded by adjacent neurons. Following this topographical organization, a retinotopic map of the available visual information is relayed by the striate cortex (V1) to areas V2, V3, V4 and MT/V5. In passing from central (or foveal) to peripheral vision, receptive fields become progressively larger. As a consequence, visual acuity degrades with eccentricity, resulting in an inhomogeneity of resolution in passing from the central to peripheral visual fields. In terms of cortical topography, the amount of cortex devoted to central vision is magnified at the expense of peripheral vision, so that the central 10 degrees of the visual field occupies roughly half of area V1 (Hubel & Wiesel, 1974). From the cell's perspective, the region of space from which visual information is processed, the RF, is highly restricted in the pre-cortical visual pathways and early visual cortex. RFs increase both in size and complexity as one ascends the visual hierarchy, enabling more complex representations of information. This produces a mismatch in FF and FB pathways given the large RF of higher-order areas projecting back to lower areas with much smaller RFs (Salin et al., 1992), implying that FB does not lead to an activation of the classical RF of early stage neurons (Angelucci & Bullier, 2003).

FF, FB and lateral information flow

In the 1970s and 80s, several research groups used anterograde (Martinez-Millan & Hollander, 1975; Spatz et al. 1970; Tigges et al., 1973) and retrograde (Rockland & Pandya, 1979; Lund et al., 1975; Fiskens et al. 1975; Wong-Riley, 1974; Kennedy & Bullier, 1985) tract-tracing to study inter-areal connectivity of early visual cortex. The emerging consensus was that FF and FB connections are frequently reciprocal with laminar-specific properties in which FF projections mainly originate from superficial layers and target layer 4, while FB pathways

mainly originate from deep layers and target any layer other than 4. Together with the ascending and descending hierarchical streams, lateral connections (intrinsic to a hierarchical level) were identified, a finding that was clearly in contradiction with the most widespread model at the time, i.e., visual perception resulting strictly from FF process (Hubel & Wiesel, 1962; 1977; 1968; 1976; Barlow, 1972).

The Dual Counter Stream

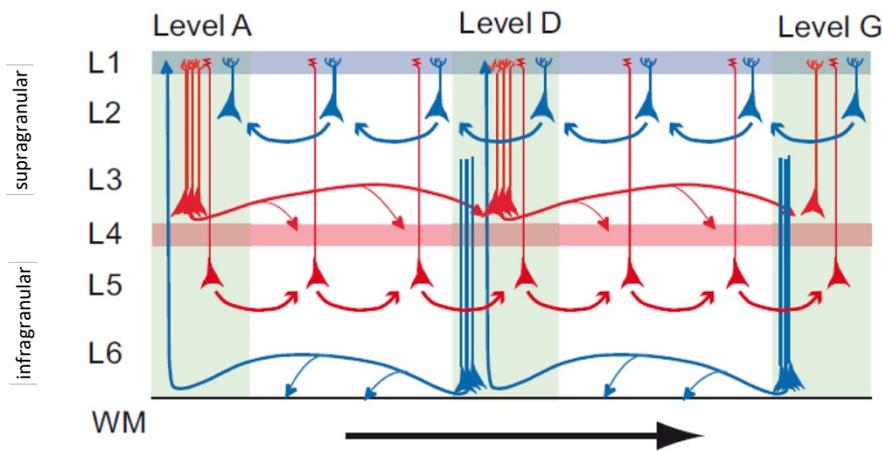


Figure 2: Anatomy of Hierarchy: The Dual Counter Stream
In red: feedforward streams; in blue: feedback streams

Markov and collaborators (2014b) showed; (i) that the supragranular FB exhibits a short distance, point-to-point connectivity while the infragranular FB is a long-distance and more diffuse connectivity; (ii) FB are twice as numerous as FF; (iii) on average FB cross more hierarchical levels than do FF; (iv) on average FF are stronger than FB. From a previous study (Markov et al., 2013a) it emerged that the cortical inter-areal matrix has a density of 66%, meaning that nearly 2/3 of the possible connections actually exist. This density is nearly twice that of the database exploited by Felleman and Van Essen (1991). The high density of the cortical matrix has important consequences for the cortical hierarchy since each area on a particular level projects to all the higher and lower areas as well as areas on the same level (Markov et al., 2013c). Consequently, the pathways ascending and descending the cortical hierarchy are massively parallel, and the seriality of earlier models is seriously eroded. Given that most of the cortical input to a given area is considerably weaker than the LGN input (which is thought to be just enough to activate an area) and while it seems reasonable to believe that the response properties of an area are consequent to its connectivity profile, it leads to the conclusion that the activation of a cortical area is a cooperative process (Markov et al., 2014a; see review Passingham et al., 2002). Furthermore, the Felleman and Van Essen database was compiled from data generated across different laboratories using very different criteria and plane of section which led to large errors in count data with respect to the consistent criteria used in the Markov et al., 2014a study.

On top of revealing a multiplicity of pathways across cortical layers, the Dual Counter Stream architecture (Fig.2) re-defines key properties of cortical projections. The convergence of cortical FF connections (i.e., “from many to few”) is now considered to be shared by the supragranular FB pathway, while the notion that FB connections are divergent (“from few to

many”) seems to be specific of the infragranular FB type (see Vezoli et al., 2021 for further discussion).

The structural hierarchy is now known to have a functional counterpart. In macaque distinct neural oscillatory patterns were observed at distinct cortical depths (Maier et al., 2010; Buffalo et al., 2011; Bosman et al., 2012). Pascale Fries and his team were able to show that these oscillatory patterns replicated the laminar connectivity distribution described in anatomy by Kennedy’s team (Markov et al., 2014b), enabling these authors to construct an extensive functional hierarchy of the macaque visual cortex (Bastos et al., 2015). In a second step, these authors used magnetoencephalography to look at oscillatory patterns in human cortex, and by reference to areal homologies to macaque cortex, were able to construct a functional hierarchy of human cortex (Michalareas et al., 2016). This was a remarkable achievement where research in non-human primates was able to play a key role in furthering our understanding of the human brain. Altogether, structural models of cortical hierarchy appear to be particularly coherent with the theory of predictive processing and the notion of the Bayesian brain (Bastos et al, 2012; Markov & Kennedy, 2013b; Feldman & Friston, 2010). Therefore, establishing a functional hierarchy in both the macaque and human brain greatly facilitates the investigation of the neural mechanisms underlying predictive processing (Friston, 2019; Shipp, 2016; Vezoli et al., 2021).

To explain the fact that each level of hierarchy comprised several areas interconnected via lateral connections, Felleman and Van Essen emphasized a preexisting view (Henry et al, 1983) that, particularly in the visual system, multiple processing pathways run in parallel.

A multiplicity of pathways

Retinal projections to LGN reflect two main streams of visual information: magnocellular (M) and parvocellular (P) pathways. These are named after the type of retinal ganglion cells they receive input from (M(agno)- for large cell bodies, P(arvo)- for small cell bodies). In terms of function, it is well established that M cells have larger receptive fields and are characterized by a peak of sensitivity for low spatial and high temporal frequencies. In other words, the M pathway responds preferentially to large and high-speed visual stimuli, hence is suitable for contributing to processing motion information. However, they are less responsive to wavelength and chromatic differences. In contrast, P cells have smaller RFs, and are sensitive to high spatial and low temporal frequencies, therefore responding preferentially to small and slow stimuli. In addition, P cells respond differentially to wavelength and are sensitive to chromatic differences, hence are implicated in processing color information. This anatomical and functional segregation is partially preserved throughout the visual hierarchy, with the M path dominating projections towards the dorsal visual stream (the “Where” path, e.g., motion representation), and the P path preferentially towards ventral stream (the “What” path, e.g., color information), with cross talk interactions occurring at different stages of the circuitry (reviewed in Previc, 1990).

We have seen that the cortical hierarchy is well defined in terms of anatomy. However, investigating functional connectivity is equally important to determine what are the functional consequences of the structural hierarchy. To answer this question, a growing number of neuroimaging and electrophysiological studies appeared in the 2000’s, aiming to distinguish top-down from bottom-up signals, as well as demonstrating mental representations that are supported across hierarchical levels.

The fact that FF and FB constitute distinct neural populations (Markov et al., 2014b) is recognized as a key requirement enabling distinct functionalities of generative top-down networks (Friston, 2018; Markov and Kennedy, 2013c; Shipp, 2016), a concept that will be described in the next part.

1.3 Towards the definition of a functional hierarchy

The importance of top-down generative network

Conceptually, bottom-up visual input aims to break down or filter high-dimensional sensory representations into low-dimensional information represented at higher levels of the hierarchy. Conceivably FB pathways invert this process by generating high-dimensional images starting from low-dimensional features (Hinton, 2007). This notion is at the core of what is referred to as a generative top-down network (for review, see Vezoli et al., 2021).

V1, a dynamic blackboard for mental representation

Investigating the earliest cortical stages of visual processing enables understanding of the gradual increase in complexity of neural representations built up from relatively simple and fundamental visual features (contrast, orientation, color, motion, etc.). Hypothetically, these attributes will correspond to the substrate of our mental imagery. Moreover, the retinotopic spatial organization of visual information renders the information easier to track. As a result, we have good insights on the connectivity patterns of early visual cortex. For example, considering V1 total input connectivity, only 1 to 2% arises from the LGN whereas around 80% originates from intrinsic connections and the remaining 18% from higher-level cortical areas (Markov et al., 2011). fMRI experiments comparing the neural correlates of visual perception, working memory and mental imagery in early visual areas (from V1 to hMT+) have shown that these three different visual modalities actually share highly similar patterns of activity, hence reflecting common neural correlates (Slotnick, 2005; Albers et al. 2013; Lawrence et al., 2018). These findings indicate that early visual cortex plays a specific role in maintaining information about recent visual history, giving access to local fine-tuned feature information. As expected, this does not seem to be the case in higher-order areas showing more complex and global information processing (Harrison & Tong, 2009). Moreover, the ability to reconstruct the mental image of an object that is absent from the visual field, incomplete or ambiguous, highlights the importance of FB signals in shaping and predicting sensory signals (Muckli et al., 2015; Papale et al., 2022). This raises strong arguments against the outdated view of FB connectivity as exerting a relatively weak and strictly modulatory influence on cortical processing.

As stated previously, it is generally accepted that V1 encodes visual information in a topographical manner, reflecting retinal input, and that this retinotopic structure is also preserved in the earliest cortical stages of visual processing (V2, V3, V4). Moreover, we have seen that these areas contain columnar functional units whose activity is maximized for a preferred feature. However, determining how this structure evolves in higher-order areas to represent more complex/naturalistic scenes is still debated among researchers. We know that the representation of information follows specific neural dynamics inherent to cortical structure, but which ones? To tackle this question, researchers initially proposed the “efficient coding” theory, the general concept of which states that sensory neurons encode maximal

information of visual input given internal constraints (metabolic cost, noise, etc.) (see review Chalk et al., 2018). Many neurocomputational models have been proposed, differing in the nature and substrate of the encoded information, the choice of constraints and the field of application. Such mathematical frameworks seek to model cortical dynamics involved in sensory and cognitive tasks used to represent sensory input. I will shortly introduce two of these models that appear particularly relevant in the context of my thesis project.

Identifying functional units

As first described in cat and macaque cortex (Hubel & Wiesel, 1962; 1968), early visual cortex shows a remarkably well-preserved functional organization between V1 and MT/V5 hierarchical levels. Functional columns and patches of varying size and spatial arrangement were characterized, based on both immunohistochemistry (e.g., Cytochrome Oxidase) and high-resolution neuroimaging (e.g., optical imaging, fMRI), by investigating neural activity in response to a specific stimulus feature. Similar structures and recurrent patterns were reported across macaque and human primates. This segmentation of feature-specific interdigitated pathways was first described in V1 with columnar units specific to orientation (“*pinwheel*”), color (“*blob*”), spatial frequency, ocular dominance and disparity (Livingstone and Hubel, 1984; Cheng et al., 2001; De Valois et al. 1982ab; Adams et al., 2007; Tsao et al., 2003; Guan et al., 2021). Subsequently, researchers extended these investigations to higher-order early visual areas (V2, V3, V4, MT/V5). They were able to define similar functional maps reciprocally connected across hierarchical levels, thereby showing evidence for parallel feature-specific visual pathways (Yoshioka & Dow, 1996; Albright, 1984; Diogo et al., 2003; Ahmed et al., 2012; Tang et al., 2018a; Malonek et al., 1994; Zimmerman et al., 2011; Schneider et al., 2019). Depending on the areas involved and the selectivity of the response for a given physical parameter, some areas appeared to be more specialized than others (e.g., V4 for contours/shapes and color, MT/V5 for motion, etc.). Taken together, these results show that distinct attributes of visual input are processed in parallel within early hierarchical stages of visual processing. Nonetheless, in the aim of reaching a more fine-grained functional map of the cortex, Li and collaborators (2019) proposed a revision of the current representation of the orientation hypercolumn by identifying distinct subunits encoding multiple elementary features of visual contours (preferred orientation, optimal spatial frequency and temporal frequency, location and size of the classical RF, surround suppression property, response latency, etc.) within a single pinwheel. This functional representation of RF properties of V1 strengthens the notion that it encodes general building blocks used for more specific representations formed in higher-order visual areas (Barlow, 1972; Li et al., 2019; Tang et al., 2018b). This is well illustrated in the temporal ventral stream where there is a progressive construction of complex shape/features representations (e.g., leading to object and face recognition) (Fujita et al., 1992; Tanaka et al., 1996; Wang et al., 1996; 1998; Tsunoda et al., 2001; Sato et al., 2009; Tsao et al., 2008).

A major research effort is being made in identifying the neural code of our internal representations, from their initial encoding to their storage. V1 stands out as the key area in receiving and filtering sensory input, but also integrating information from higher-level neural representations, and for this reason remains the focus of interest of a large number of studies.

Efficient coding

The efficient coding theory, as first introduced by Horace Barlow in 1961, hypothesizes that sensory coding (as assessed by sequences of action potentials of neurons) in the brain follows a specific coding scheme aiming to encode accurate input information while minimizing the associated metabolic cost. Indeed, propagating spikes throughout the brain is energy-demanding and seems to be strategically optimized by limiting the neuronal encoding to the strict minimum while ensuring sufficient signal-to-noise ratio. It is important to recognize that the standard definition of efficient coding fails in describing precisely what network dynamics could support the efficient coding of complex sensory information given the internal constraint of the physical network in place. Moreover, addressing the question of the energy minimization appears fundamental (Johnson et al., 2016). To this aim, researchers proposed that independent and local neural systems extract and encode information following a “sparse latent structure”. This led to the sparse coding theory, subdividing neural networks into small neural units firing synchronously and sparsely at a given time. Originally, this was particularly adapted to describe orientation and motion selectivity in the early visual cortex (Olshausen & Field, 1996; Bell & Sejnowski, 1997; van Hateren & van der Schaaf, 1998; van Hateren & Ruderman, 1998) and was inspired from Barlow’s observations (1972) who suggested that any given information was encoded by fewer and fewer neurons as one moves up into the visual cortical hierarchy. Since then, many lines of evidence have converged, supporting and developing the computational strategy underlying sparse coding (Olshausen & Field, 2004; Rolls & Tovee, 1995; Vinje & Gallant, 2000). This type of neural computation presents several advantages. First, it optimizes memory storage capacity by maximizing local learning associations in highly specific neural networks. Second, it makes the statistical signature of signals more explicit, allowing the representation of increasingly complex visual input at subsequent levels of processing. Finally, it allows a reduction of metabolic cost by limiting the number of cells that fire simultaneously. Numerous electrophysiological recordings suggest that sparse coding is a ubiquitous strategy that applies to various sensory modalities and across different organisms (Brecht et al., 2003; Tang et al., 2018b).

Predictive processing theory

However, sparse coding theory does not resolve how neural systems define and prioritize context-relevant information. Bialek and coworkers (Bialek et al., 2006; Palmer et al., 2015) adjusted the model, by positing that efficient coding aims to maximize any past information that contains predictive value about the future. As a consequence, neurons would preferentially encode what deviates from statistical regularities. This coevolved with what we now refer to as predictive coding or processing theory, which is a descendent of von Helmholtz’s unconscious inference (1867). This form of efficient coding posits that while top-down generative networks encode predictions about future events (implementing Bayesian inference), bottom-up prediction error signals are generated to estimate and update prior predictions in the light of sensory input (Rao & Ballard, 1999; Clark, 2013; Hohwy, 2013; de Lange et al., 2018; for review see Vezoli et al., 2021; Maric & Domijan, 2022). Although the idea that neural systems make inference about sensory input using internal or generative models is quite common, the notion of predictive processing is controversial. Nevertheless, this theory, is one of the most neurobiological plausible candidates for exploring the concept of brain inferences (Srinivasan et al., 1982; Mumford, 1992; Rao & Ballard, 1999; Friston, 2010).

The canonical microcircuitry and the issue of cortical activation

The canonical cortical microcircuit theory explores the electrophysiological properties of the local cortical connectivity, notably with respect to within-laminar connections (note that local and microcircuit are used interchangeably). In vivo intracellular recordings made in area V1 in cat cortex revealed that thalamic input to area V1 was greatly amplified locally by recurrent excitatory loops within layer 4 of area V1 (Douglas et al., 1995). Subsequent quantification of the local anatomical cortical connectivity showed that the recurrent excitation is a highly characteristic feature of the local circuit particularly for the thalamic recipient layers (layer 4 and layer 2/3), but less of the infragranular layers (Binzegger et al., 2009). Although excitatory neurons outnumber inhibitory neurons by a factor of 5, such a computational framework explains how the inhibitory signal is amplified locally, hence exerting a control on the balance between excitatory and inhibitory signals in order to optimize the RF response (Douglas et al., 1989; 1991; 1995).

The original study describing the canonical microcircuit theory showed that amplification of the weak input to neurons in layer 4 was responsible for the orientation response. This idea challenged the Hubel and Wiesel theory that the orientation response was the consequence of the convergence of LGN to area V1, a proposition for which empirical evidence was provided by Ferster et al (1996), who claimed that the role of the recurrent excitatory loops was limited to the amplification of signal, not its generation. However, recent work in mouse visual cortex using modern techniques shows a non-random connectivity of cells in the area V1 layer that receives thalamic input. These authors describe a lognormal distribution of synaptic weights where strong connectivity was shown to be restricted to neurons sharing similar RF properties (Cossell et al., 2015). These findings suggest that each neuron is generally functionally connected to neurons with similar response properties, thereby ensuring a selective signal amplification. This idea is highly relevant to cortico-cortical inter-areal connections as the vast majority of these connections are much weaker than LGN cortical input and illuminate the observation of Tsodyks and collaborators (1999) that ongoing (i.e., spontaneous) activity in the cortex reflects the backbone of local strong connections.

Recently, a link was made between the canonical cortical microcircuit and its application to the amplification of FF and FB inputs to a given area (Vezoli et al., 2021). According to this view, local amplification of incoming FF signal supports the function of shaping the RF of the target area (Hubel, 1995). Basically, FF processing is the success story of neuroscience over the last 50 years. This contrasts with the FB pathway which remains to this day poorly understood but is thought to be relaying contextual information to lower-level areas (Gilbert and Li, 2013; Zipser et al., 1996 ; Bullier, 2006) (see further discussion of FB pathways below in *Identifying feedback functions*). The importance of the amplification of input signals to the cortex suggests that *all* inputs to the cortex have a modulatory function, therefore blurring the distinction of modulatory and driving inputs (Vezoli et al., 2021)

The Bayesian brain

Bayesian inference is a method of statistical inference (derived from Bayes' theorem) that consists in updating the probability for a prior hypothesis as more evidence/information becomes available.

According to Bayes theorem:

$$P_{post} = P_{prior} \cdot L$$

Where P_{post} is the posterior conditional probability distribution, P_{prior} the prior probability distribution (reflecting prior belief and the uncertainty that it involves), L the Likelihood, a density function defining the probability distribution relative to a specified statistical model (integrating new evidence/information)

Applied to sensory processing, Bayesian inference assumes that the brain is an allostatic system, which naturally tends to minimize entropy (i.e., error), or the level of surprise associated with the acquisition of new sensory evidence, as represented in the likelihood. It is suggested that the Bayesian brain maximizes the likelihood (compatibility) of the model by updating the probability for the prior hypothesis while taking into account new evidence. This probabilistic model is particularly relevant to explain how the brain can handle incomplete, noisy, or ambiguous sensory input signals (Friston, 2012). As highlighted by the free energy principle (Friston, 2008; 2010), prediction error is undesirable because it signals that the brain has lost allostatic balance with the environment. The free energy principle also succeeds in describing how to fit generative models to sensory signals following a hierarchical structure where predictions and prediction errors flow across multiple processing stages (Friston, 2008; 2010).

From this perspective, the Bayesian approach can be regarded as a way to filter and interpret the massive and noisy flow of sensory information. This idea constitutes the foundation of predictive processing theory, assuming that brain machinery encodes and updates its internal representations by estimating the most likely explanation considering “what is already known” (i.e., prediction) and new relevant information that needs to be integrated (i.e., prediction error) (see reviews de Lange et al., 2018; Yon et al., 2019). Therefore, the prediction error signal (bottom-up) leads to updating the prior representation (top-down prediction) by indicating when sensory input does not meet expectations. A third parameter to consider is the intrinsic precision weighting signals which modulate the influence of the prediction error signal. For example, a high-precision weighting will enhance the influence of the prediction error on the posterior predictions, shifting it “away from priors”, whereas a low-precision weighting will have the opposite effect.

Although these theoretical models of efficient coding can appear complementary on some aspects, it is important to point out that they are constrained by different parameters (information saliency, energy cost, neural network, etc.). Moreover, an important source of confusion is caused by the fact that many publications attribute the same name to models using very different mathematical formulas. For this reason, defining a unifying framework that enables comparing and/or combining distinct types of processing is extremely informative (Chalk et al., 2018).

Identifying feedback functions

It is generally accepted that the main function of FF connectivity in perception is to convey sensory signals ascending the cortical hierarchy while expanding receptive fields both in size and complexity. According to predictive processing theory, the amount of sensory signal ascending to the brain is proportional to the level of surprise. Therefore, the more visual input deviates from predictions, the stronger FF signal and inversely, the more visual input conforms to the prediction the weaker it becomes. However, there is no clear explanation of how the brain constrains the bidirectional flow of information between FF sensory input and FB predictions. It was argued that this predictive shaping of information confers a direct top-down shaping of visual processing called “cognitive penetrability of vision” (Lupyan, 2015, 2017; O’Callaghan et al. 2017; Hohwy, 2017). Moreover, according to some authors, the modulation would differ between early and later stages of processing (Raftopoulos & Zeimbekis, 2015; Raftopoulos 2009; 2014; 2019; for review see Maric & Domijan, 2022). Obviously, FB function still raises debate and is the object of investigations in a broad range of fields. For example, transcranial magnetic stimulation in humans showed that fast FB projections from MT/V5 to area V1 are necessary for visual awareness (Pascual-Leone & Walsh, 2001; Tong, 2013). Among the most studied implications of FB signals, a few examples are shaping mental representations (e.g., working memory, van Kerkoerle et al., 2017, Lawrence et al., 2018 ; mental imagery, Slotnick et al., 2005; Harrison et al, 2009; Albers et al, 2013), building perceptual inference based on prior knowledge (i.e., expectations) (Summerfield et al., 2008; 2011 ; Rahnev et al., 2011) and attentional modulation (i.e., behavioral relevancy of available information) (van Kerkoerle et al., 2017), as well as resolving local ambiguities (Kok et al., 2014; 2016; van Kemenade et al., 2022). In general, all of these functions have in common to provide guidance to the visual system.

2. Investigating FB function through perceptual integration

The first parameter of importance when aiming to study Blood-Oxygen-Level-Dependent (BOLD) FB signals independently from FF activity lies in the choice of the experimental paradigm. In order to modulate FB activity while keeping FF signal constant, I designed a stimulus that would trigger spontaneous perceptual switches (in the absence of changes in stimulus physical properties).

Extracting visual information from a general context and shaping mental representations has been long thought to engage visual cortical areas exclusively. However, the emergence of studies investigating multistable perception at the end of the 20th century (see review Leopold & Logothetis, 1999) marked a turning point in the understanding of local/global representations. Notably, the fact that associative areas play an essential role in the integration of visual and non-visual information highlighted the influence of top-down signals in perceptual integration.

Integrating spatial cues

Building a representation of three-dimensional visual space involves exploiting spatial cues that can be derived from various features such as binocular disparity, shading, shape, etc. as well as using prior knowledge to convert the retinal information including disparity of visual elements in both retinas into depth information. Therefore, perceptually shaping 3D mental

images out of 2D retinal input requires selecting, integrating and translating spatial cues into coherent and robust global interpretations of a visual scene. Sensory signals are by essence ambiguous, associated with high uncertainty, and it is assumed that the Bayesian brain tries to explain or interpret these by fitting the most likely prediction and minimizing errors. Different strategies have been described that could disambiguate visual input. First of all, by multiplying the sources of information that are context-relevant, thereby more information is provided to the brain to make inferences about the sources of retinal stimulation. Second to the massive collection of data, comes the objective of selecting the most reliable perceptual estimates, that is the one that shows minimal noise. Prior experience will help constrain the range of possibilities to the most likely interpretations. In the objective of reducing noise that is inherent to any type of sensory signal, the perceptual estimate with the lowest variance is selected under the Maximum Likelihood Estimate (MLE) rule (i.e., method estimating the parameters of an assumed probability distribution, given some observed data). In naturalistic scenes, combinations of sensory signals benefit from multimodal cooperation. However, in the next sections I will review a few examples of purely visual integration.

Illusory contours and surface filling

Visual illusions generate vivid sensations of features which are not physically present. In the field of research, illusory stimuli are widely used when investigating perceptual brain inferences (Brown & Friston, 2012). One well-known example of this type of stimulus is the Kanizsa illusion which is traditionally composed of several “pac-man” inducers spatially arranged in a way to suggest an overlaying surface (Fig. 3(A)). The configuration triggers the perception of illusory contours, thereby revealing an illusory figure that appears brighter than the background; this effect is referred to as surface filling-in. The same type of phenomenon occurs in figure/ground segmentation of a textured surface in which a small region shows texture discontinuities/contours that are detected through boundary completion, making the delimited figure pop-out from the background. It is known that FB from higher areas to V1 has a unique function in surface filling-in (whereas boundary detection would be more automatic and already processed at the level of V1, see Poort et al., 2012). Moreover, laminar neuroimaging in humans and invasive recording in macaque suggest that different FB pathways might be involved in boundary completion (superficial layers FB) surface filling-in (deep layers FB) (Kok et al., 2016; Self et al., 2013). Another example of long-range filling-in that generates color surfaces signaled by distance chromatic contours and that appears to involve feedback in multiple pathways is the Watercolor Effect (Pinna et al., 2003; Gerardin et al., 2018; Devinck et al., 2019). Evidence for multiple feedback signals have been presented as well for percepts related to “real” and “imagined” visual experiences (Bergmann et al., 2019 bioRxiv).

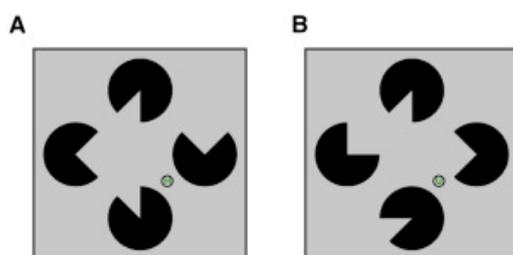


Figure 3: Example of Kanizsa stimulus (Kok et al., 2016)
 (A) Surface filling-in effect
 (B) No effect

Local vs Global perceptual integration

Generally, perceptual integration occurs when local features are integrated into a global percept. This phenomenon can be experienced with any physical parameter such as contrast, color, orientation, shape, motion, etc. Encoding visual input into internal representations is by nature ambiguous, especially in low-level areas, but this ambiguity generally tends to decrease in higher-order areas, as more complete “ensemble representations” are built. However, when these ensemble representations of local features remain ambiguous, giving rise to several equiprobable interpretations, it can create a multi-stable visual perception, that is a spontaneous change of perceptual state, given unchanged physical stimulation. In the particular case where this type of ambiguous visual stimulus is experienced over a long period, it can trigger successive exchange between two (or more) equiprobable and mutually exclusive perceptual states: perception is then considered as being bistable (or multi-stable). The dynamics of bistable perception will be detailed more thoroughly in the second part of this introduction.

Perceptual functions based on mental representations, multi-modal integration, and inference illustrate the influence of top-down activity on early visual processing. The model of motion integration is a good candidate to study feedback function because of the relatively strong and direct anatomical FB connectivity from motion-selective areas MT/V5 to V1, two hierarchically and physically distant cortical areas. Based on the central role of MT in motion processing, and more specifically motion integration, this suggested a key role for this FB pathway in generating predictions about pattern motion information reflected as early as area V1. Although motion detection systems in area V1 are presumed to only respond to local “components” (or 1D) motions (e.g., gratings of a narrow band of orientations), while motion integration of “pattern” is processed in higher-order areas in the hierarchy (such as the motion complex hMT+), several electrophysiological studies have reported evidence of pattern-selective responses as early as area V1 in awake macaques (Pack & Born, 2001; Guo et al., 2004). Supposing that the cells in area V1 are narrowly tuned to orientation and have small RFs, this suggests that the fact that area V1 exhibits activity in response to complex motion signals reflects higher-level computations. Bistable motion paradigms enable tackling this question by focusing on internally generated changes of perception. Under the right conditions, the phenomenon generates two widely different percepts (“component” or “pattern” representation of motion) that can be dynamically reported by the observer and serve to define the experimental conditions.

The dynamics of bistable perception will be developed in the next section, showing how the integration of motion signals, a model that has been extensively studied, provides a direct illustration of the idea of hierarchical information processing.

3. A well-known model of visual hierarchy: Motion integration

3.1 Motion signals

Definition

A motion signal is defined as a series of images presented sequentially and showing spatial correlations. Visual motion signals are composed of multiple features that the visual system must extract over space and time in order to encode or to detect movement. Neurocomputational models representing motion processing in the visual system describe

multiple pathways along the visual hierarchy, each dedicated to specific aspects of motion signals (Snowden & Hammett et al., 1992; Ledgeway & Smith, 1994; Lorenceau, 2001; Sun et al., 2014; 2015).

I will take an extensively studied example of a one-dimensional motion stimulus, a grating (defined as a periodic repetition of oriented black and white lines), as illustration.

First-order: Fourier motion energy

The Fourier transform is a mathematical tool widely used to decompose complex signals (which are not necessarily periodic) into the sum of periodic functions of different frequencies and phases. An advantage of such a transformation is the ability to represent a complex signal in terms of simpler components. The distribution of the amplitudes and phases of the transform as a function of frequency constitutes the spectrum of the signal. For periodic stimuli, the Fourier decomposition yields a discrete series of components whose sum reconstructs the original signal. The fundamental frequency corresponds to the lowest frequency (i.e., with the longest period) and defines the periodicity of the signal. Any multiple of this fundamental frequency is called a harmonic, and together with the fundamental, they constitute the frequency spectrum of the periodic signal.

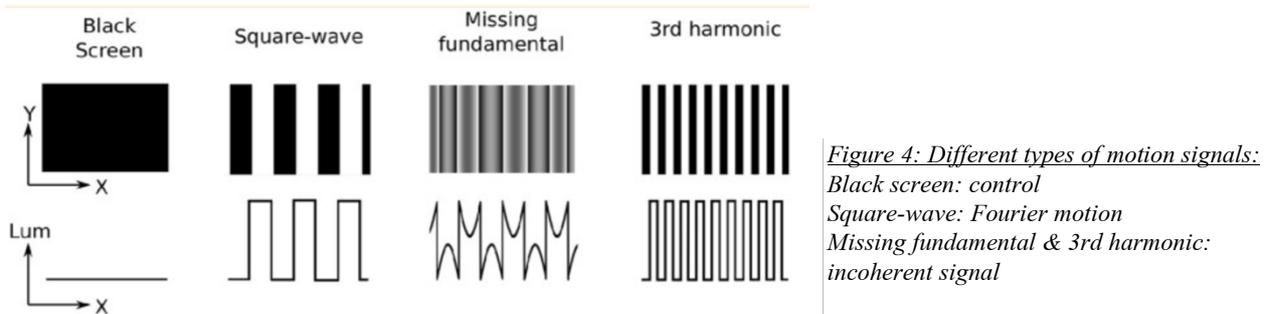
Classic theories of motion perception assert that, first-order motion signals are extracted by “elementary motion detectors” presumed to be mainly located mainly in V1 (first described by Reichardt, in 1957, see also Adelson and Bergen, 1985, Van Santen and Sperling, 1984; 1985). This motion detection system would extract “Fourier energy”, which is proposed to integrate luminance as a linear function of luminance over time (Adelson and Bergen, 1985; Van Santen and Sperling, 1984; 1985; Watson and Ahumada, 1985; Chubb and Sperling, 1988, Emerson et al., 1992, Lu and Sperling, 1995; 1999; 2001). This proposition is supported by the observation that initial eye movements are invariably made in the direction of the Fourier energy (Chen et al., 2005; Montagnini et al., 2007)).

Nonetheless, although first-order motion processing has the advantage of being fast (generating response within 60 ms approximately), it does not provide sufficient information to describe all types of motion. For example, it is possible to induce perceived motion by reversing the stimulus contrast over time (flickering target/background) or by using complex textures. However, in these examples, first-order (or luminance-based) motion detection will not be effective if the average luminance remains unchanged over time.

Second-order: Non-Fourier motion energy

We have just seen that some stimuli require more complex (“non-Fourier”) computations of the motion signal, involving higher-order detectors. Such motion-selective cells can derive non-linear motion energy from luminance-based features such as contrast, boundaries, texture, etc. (Ramachandran et al., 1973; Chubb & Sperling, 1988; Chubb et al., 1994; Baker, 1999). It is presumed that non-Fourier motion detectors are more likely to be found among higher-order visual areas (e.g., MT/V5) and are absent or rare in V1. According to Wilson et al. (1992), both Fourier/non-Fourier signals involve distinct pathways along the V1-V2-hMT+ axis, providing complementary information. They both contribute to identify the velocity (i.e., speed and direction) of the seen motion. One hypothesis is that both Fourier/Non-Fourier motion energies

would eventually be summed into a global output signal at the level of hMT+. However, in the case of noisy or ambiguous perception, concurrent first-order and higher-order detection systems might predict different, incoherent velocities. Using Fourier decomposition, it is possible to create a stimulus that has a Fourier motion energy moving in the opposite direction to that of the other features (edges, contrast, textures, etc.) (Adelson & Bergen, 1985; Chen et al., 2005; Duchemin et al., 2022) (Fig. 4).



In an attempt to explain what happens when disambiguating motion information, or the fact that one perceptual state becomes predominant, it was proposed that a suppressive hMT+-to-V1 FB signal would silence concurrent neural units (Wilson et al., 1992; Berzhanskaya et al., 2007).

Third-order: Feature-based motion energy

These first- and second-order motion mechanisms are primarily monocular (i.e., specific to one eye). Nonetheless, the existence of purely binocular (i.e., relative to both eyes), and in general more global motion detection system, has promoted conjectures of a third-order mechanism. This global motion detection system would be specific to stimuli that produce apparent motion, that is motion perceived only if the observer selectively attends to one of the embedded “features” (e.g., edge, pattern, shape, etc.). Importantly, such features would by definition be invisible to first- and second-order motion systems. Hence, they require a more global processing, that is made possible by integrating motion signals into a salience map (i.e., a representation of what is the most salient information to the observer in the image). As a result, the direction of the apparent motion differs depending on the focus of attention (Lu & Sperling, 1995). Experimentally, such feature-selective mechanisms were investigated by designing ambiguous moving gratings in which the Fourier motion energy dominates during the earliest stage of the stimulus presentation, before rapidly switching to a feature-selective dominant motion energy in a different direction (e.g., Georgeson & Harris, 1990; Smith, 1994).

Such feature-selective mechanisms are involved in several well-studied experiments involving moving gratings and where the perceived motion is not perpendicular to the grating’s orientation (contrarily to what Fourier motion energy predicts). An extensively studied example is the Barber Pole illusion, where an oblique grating moving orthogonally to its orientation (Fourier motion energy direction) is perceived as moving vertically when presented through a thin rectangular vertical aperture (Sun et al., 2015). In this stimulus, the most salient feature detected is induced by motion signals at the edges of the aperture (an end-stopping property).

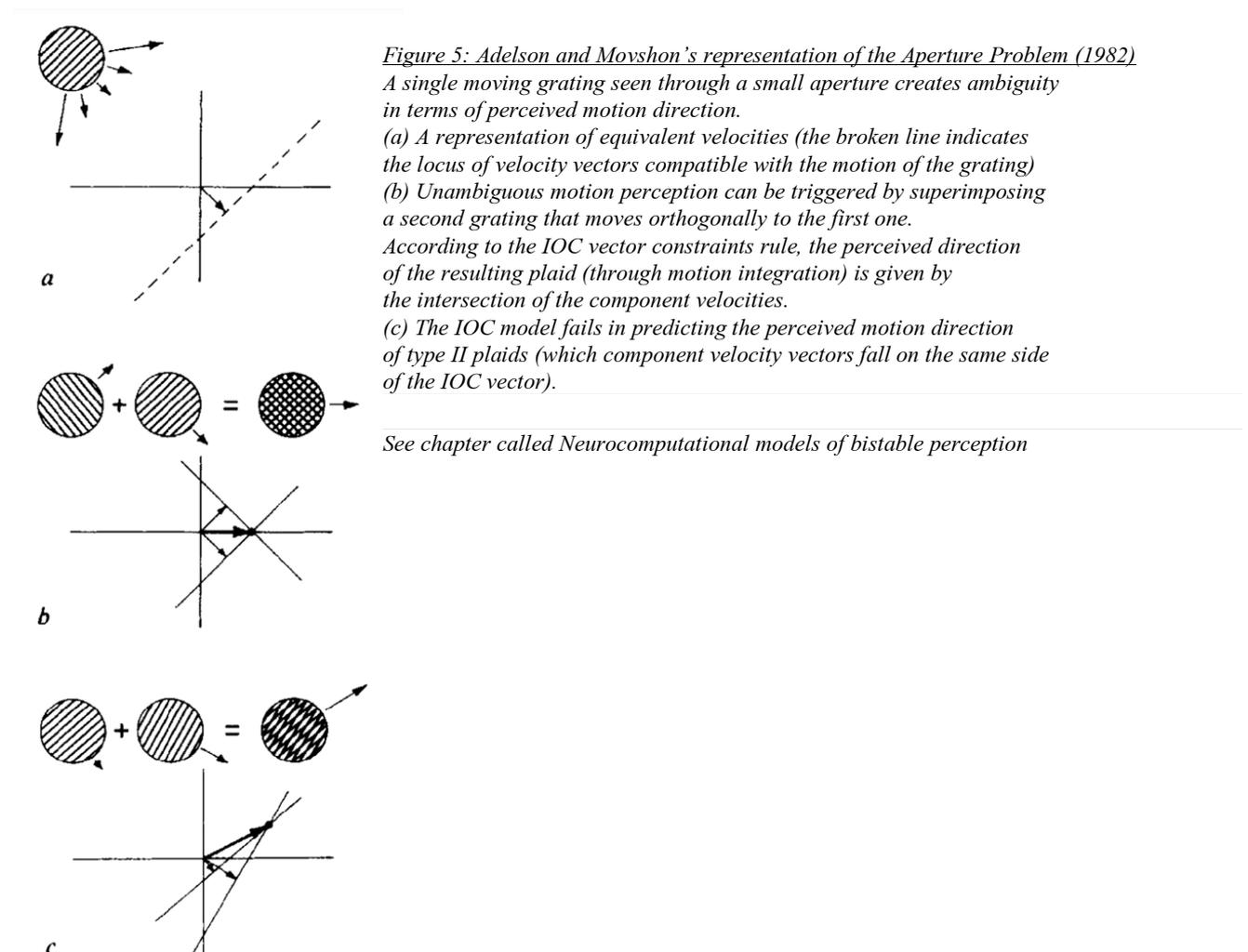
Another example involving a “small aperture” presentation is the ambiguous plaid (made of two superimposed grating components moving obliquely), that is alternatively perceived as component motions (i.e., moving gratings) or pattern motion (i.e., a plaid plaid moving in a different direction) (Adelson & Movshon, 1982; Stoner & Albright, 1992b). One hypothesis that would explain how component motions are fused into a pattern is that the sum of the two gratings results in a highly salient feature specific to the plaid, “blobs”. The blob is a structure derived from interference of luminance signals at the intersections of the component gratings. This phenomenon will be covered in a later section, *Neurocomputational models of bistable plaid motion*, in which I will describe relevant neurocomputational models proposed to explain how perceptual decisions evolve over time based on specific spatio-temporal cues.

3.2 Moving plaids and The Aperture Problem

Ambiguous infinite uni-directional (1D) motion

When viewing an oriented grating moving through a small circular aperture, in such a way that its edges are masked, i.e., so that the grating can be displaced in different directions of motion for a constant orientation, observers perceive only a single direction of motion, that is perpendicular to the orientation of the grating. A proposed explanation of this is that masking the endpoints of the grating renders motion signals infinite along the orientation of the lines. Any variation of motion collinear with the orientation of the grating cannot be detected because the luminance-derived signal is constant. Instead, the only measurable motion signal is along the axis orthogonal to the grating’s orientation (corresponding to the Fourier motion energy), thus explaining that we always perceive the grating as moving orthogonally to its orientation. In other words, it seems that when the available motion signal is not sufficient to discriminate between multiple combinations of direction and speed, the dominant percept is driven by first-order motion detection. In 1982, Adelson & Movshon conducted a series of psychophysical experiments on this phenomenon which they referred to as the Aperture Problem.

Resolving the Aperture Problem in the velocity space



In order to understand and represent the mechanisms underlying the ambiguity that occurs in the Aperture Problem, we shall first consider this infinite moving grating presented at a given velocity (i.e., speed and direction). Such motion can be represented in *velocity space* by a vector whose length and angle, respectively, express its speed and direction (i.e., the velocity vector). All of the possible interpretations of the direction and velocity of the grating when viewed through the aperture fall along a line perpendicular to the motion vector shown by the dashed line in Figure 5a.

Adelson and Movshon proposed that in order to resolve the Aperture Problem, it requires adding a second spatial dimension, turning the infinite uni-directional motion into a 2D signal (Fig. 5b). For this purpose, they retained the concept of the moving object viewed through a limited aperture, but this time they superimposed two non-colinear moving gratings, forming a plaid, thereby providing a non-ambiguous motion signal. The motion of the plaid was perceived as different from each of its components but perceived in the direction according to the intersection of the two lines perpendicular to the component velocity vectors (Fig 5b). This was called the Intersection of Constraints (IOC) rule. This general rule was not accurate at predicting perceived motion of a different type of plaids forming a narrow angle between gratings: type II plaids (i.e., whose component velocity vectors fall on the same side of the IOC velocity vector) (Fig. 5c).

Importantly, they found that under certain conditions, the moving plaid stimulus induced two different and alternative motion percepts. Either observers perceived simultaneously the two component motions, each moving in the direction orthogonal to its orientation (i.e., Fourier motion energies) or the single direction of motion of the plaid pattern (i.e., feature-based motion energy). This competition between two main perceptual states is a direct illustration of the current hypothesis stating that several motion systems coexist, involving pathways between V1, V2 and MT.

Evidence of Component- and Pattern selective cells in visual cortex

Movshon, Adelson, Gizzi and Newsome (1985) presented the first evidence supporting the two-stage feedforward model, thereby explaining the neural basis for the integration of oriented moving components to generate a response to the motion of the plaid pattern. They studied direction-selectivity using single-cell recordings in V1 and MT/V5 of both cat and macaque while characterizing specific direction-tuning responses to moving gratings and plaids (Fig. 6(A)(B)). Since then, a number of studies have replicated their results, showing evidence of a distribution of Component- and Pattern-selective neurons that is related to the hierarchical level of visual cortical areas (see Fig. 6(C) by Khawaja et al, 2009; Rodman & Albright, 1987; Wallisch & Movshon, 2019). On the one hand, “Component” cells exhibit narrow direction tuning curves, highly specific to the component motion directions (i.e., orthogonal to the orientation of each component grating). When presented with a moving plaid, such cells exhibit a bi-lobed or “rabbit ears” profile with two distinct peaks of activity in response to the motion direction of each of the component gratings independently. On the other hand, “Pattern” cells show a uni-modal tuning curve that preferentially encode the integrated motion direction of the two superimposed moving component gratings. Interestingly, a “Pattern” cell will produce a similar response to a single moving grating as that of a plaid pattern for equivalent global motion direction.

Most of the studies on motion integration were performed in anesthetized animals. The findings that V1 cells were selective to grating components (i.e., bi-lobed tuning curve in response to a plaid) and MT/V5 cells responsive to plaid motion (unimodal tuning to a plaid) supported the two-stage feedforward model, in which component sensitivity is restricted to early visual areas and pattern sensitivity emerges in higher-order cortical areas. However, subsequent reports in awake and behaving animals found evidence for pattern sensitivity in cells as early as area V1 in macaque (Pack & Born, 2001; Guo et al., 2004; Khawaja et al., 2009). One possibility would be that V1 cells combine their signals to generate pattern selective responses, but this seems unlikely given the limited size of V1 RFs and the spatially limited range of intrinsic connectivity in area V1. Markov and collaborators (2011) reported that 80% of back labeled intrinsic neurons in V1 are contained within 1.2 mm around the injection site, 95% within 1.9 mm. An alternative is that such responses arise from the strong feedback connections from areas MT/V5 and MST (Medial Superior Temporal) to earlier visual cortical areas (Markov et al., 2014b). We hypothesize that such feedback connectivity could be involved in the bistable perception of moving plaids.

As previously described, the V1 cells that respond only to the component gratings have narrow orientation tuning and small RFs. While integrating the spatio-temporal variation of the stimulus over its receptive field, such cells generate a stream of action potentials that encodes a one-dimensional motion signal. Trade-offs between different orientations, contrasts and velocities can generate the same output, making the identification of the motion direction of

the stimulus input ambiguous on the basis of the stream of activity of a single cell. The MT/V5 cells that respond to plaid patterns must combine the outputs of direction-selective cells responsive to different orientations in order to respond to the multidimensional direction of the plaid.

Importantly, the neural ambiguity of motion direction at the neuronal level must be distinguished from the ambiguity described above concerning the perceived motion of a grating viewed through a limiting window. Specifically, the perceptual significance of such neuronal activity is unknown. One can wonder why, perceptually, in the case of the aperture phenomenon (see section “3.2 Moving plaids and The Aperture Problem”), one perceives a monostable direction of motion rather than experiencing a stimulus whose speed and direction spontaneously shift between any possible combination. One possibility is that the visual system has built-in biases or priors to infer a particular speed and direction of motion. In the case of the moving vignetted grating, humans seem to prefer interpreting the stimulus as the one with the slowest velocity, that is the shortest velocity vector with respect to the line of equivalent velocities and whose direction is perpendicular to the orientation of the grating (Fig 5A) (Weiss & Adelson, 1998).

An alternative explanation might depend on that multiple single cells with different but overlapping orientation tuning curves all respond to the grating in the window, i.e., to the extent that their tuning curves are sensitive to the spatio-temporal properties of the stimulus. Perhaps, the perceived direction corresponds to the global intersection of the equivalent velocity lines for this ensemble of cells, yielding a global intersection of constraints solution. This explanation is not necessarily at odds with a Bayesian interpretation, in the sense that the spatio-temporal tuning of an individual cell can be considered a built-in prior. The distribution of responses from the differently tuned cells to a single moving grating could then function like a posterior probability distribution to assign the most likely cause, i.e., the motion direction and velocity, as the one that would generate the peak activity of this distribution, or if an IOC solution, the direction and velocity interpretation common to the ensemble of cells.

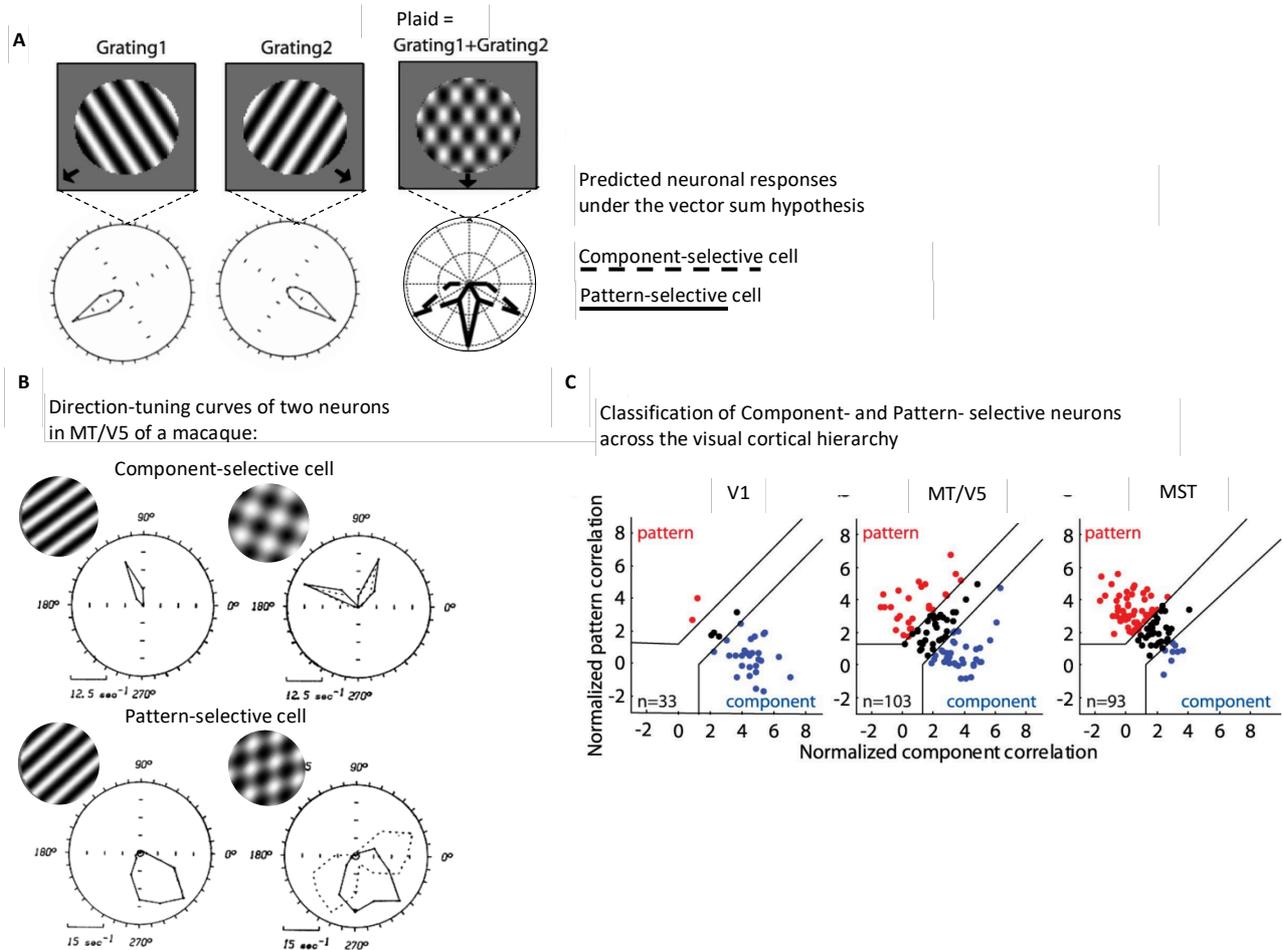


Figure 6: Electrophysiology data of component- and -pattern selective neurons at different hierarchical levels of visual cortical hierarchy

(A) Moving single gratings and superimposed pair of gratings induce different responses in neurons with different direction tuning curves as predicted by the vector sum model

(B) Electrophysiological single-cell recordings in two neurons of MT/V5 complex of an anesthetized macaque with different direction-selectivity properties.

A “Component” cell responds to the component gratings independently, and this even when superimposed (as shown by the “rabbit ears” profile of activity to a moving plaid). A “Pattern” cell responds similarly to a single grating or to the integrated (weighted sum) of the two components as long as they result in the same direction (as shown by the broader direction-tuning curve preferentially pointing towards the sum of the two component directions in response to a moving plaid) (Movshon et al., 1985)

(C) Classification analyses of Component- vs Pattern- selective cells show a gradual increase in pattern-selectivity as ascending visual cortical hierarchy (V1; MT/V5; MST: Medial Superior Temporal area) (Khawaja et al., 2009)

Note that MT/V5 + MST in macaque is equivalent to hMT+ in human

The two-stage FF model does not account for the bistable perception of a moving plaid. The duality between Component- and Pattern-selective motion detectors and the fact that their relative proportion reveals an increasing complexity ascending the visual cortical hierarchy raise the question of how we can perceptually decide on what we see, notably when visual input is fundamentally ambiguous.

3.3 Perceptual competition: from general to specific rules

Many different types of visual illusions have been studied in order to understand the critical features implicated in producing ambiguity. Binocular rivalry, which occurs when both eyes receive different information which cannot be fused, has long dominated the research field of visual competition. It has the advantage that it allows independent isolation and manipulation of the visual input to each eye, therefore establishing a direct causal relation between the two concurrent signals underlying the perceptual rivalry. Although monocular bistable perception under monocular conditions does not allow such a straightforward control of the input signal, it offers other advantages such as triggering perceptual alternation for a stimulus whose parameters are held constant. Well-known examples of bistable perception can be demonstrated with depth perception (e.g., Necker cube, 1832; Schroeder's stairs, 1858; depth/structure from motion, Wallach & O'Connell, 1953), illusory contours (e.g., Kanizsa shapes, Kanizsa, 1979), figure/ground illusions (Rubin's vase/face, 1915), motion illusions (von Schiller's apparent motion quartet, 1933; bistable moving plaid (Wallach, 1935; English translation in Wuerger, Shapley, & Rubin, 1996), etc. In principle, investigating how ambiguous sensory signals are processed by the visual system to interpret information constitutes an effective method for studying feedback predictive signals. The feedforward signal is constant and therefore, it can be argued that the perceptual transitions are triggered by an internal process, potentially of a feedback nature.

Stimulus strength and perceptual alternation

In this review, I refer to the example of monocular bistable perception of an ambiguously moving superposed pair of moving gratings (i.e., the paradigm used in my thesis project) in which perception alternates between two mutually exclusive states (component versus pattern), while both eyes receive the same input. Ideally, in the optimal conditions, the two states alternate over time with approximate equal probability. Several studies have measured the distribution of durations of the competing percepts to determine the relative strength of one perceptual state over the other, which might be referred to as the degree of bistability. The aim is generally to study the physical parameters that control the dominance of one perceptual state over the other in order to identify the conditions that lead to equiprobability between the states.

Levelt's propositions (1968) are very general rules that can apply to many types of perceptual competitions (Logothetis et al., 1996). Originally, Levelt's rules mostly described the psychophysical relation between stimulus strength and binocular rivalry (Von Helmholtz, 1867; Breese, 1899; 1909; Levelt, 1968; Blake & Fox, 1974; Logothetis & Schall, 1989; Leopold & Logothetis, 1996). Although monocular and binocular perception rely on different pathways at early stages of processing, it has been shown that the four principles known to rule binocular rivalry (Levelt's proposals) are also pertinent for monocular bistable perception (after some minor modifications, Hupé et al., 2019; Klink et al., 2008; Hupé & Rubin, 2005).

Levelt's propositions (adapted to bistable perception)

A revision of the four original proposals was experimentally tested and published by Hupé et al. (2019) to specifically adapt to the bistable perception of ambiguous moving plaids:

Proposition 1: Increasing the stimulus strength in favor of one percept increases the predominance of this percept (in terms of relative time it is perceived). In bistable plaids, it was shown that varying depth order, the angle between component gratings or their relative speed affects perceptual dominance.

Propositions 2&3: Manipulating the strength of one percept (therefore the relation of dominance) influences to a greater extent the average dominance duration of the dominant percept (proposition 2, Moreno-Bote et al., 2010 Hupé & Rubin, 2005) and the rate of perceptual switching (proposition 3). Hence, favoring one of the two percepts (e.g., either the plaid or the components) extends its own average duration (but not necessarily the non-dominant one), and decreases the rate of perceptual switching.

Proposition 4: Increasing the global stimulus strength of the stimulus increases the rate of perceptual switching. This means that a bistable plaid made of two highly salient grating components (e.g., either both high-speed or high-contrast) will maximize the rate of perceptual switching (hence resulting in very short percepts).

It is important to recall that although, based on psychophysical dynamics, monocular and binocular rivalry seem to be subject to similar stimulus constraints, both types of alternating perception are mediated by fundamentally different neural mechanisms.

3.4 Neurocomputational models of bistable perception

In this section, I review alternative models hypothesizing how bistable motion is processed by early visual cortex. Several neuromechanistic models represent competitive and modulatory interactions between cortical neural units that encode distinct representations of motion.

The Intersection-Of-Constraints model

Plaid motion has become a classical paradigm for investigating motion perception, as it facilitates the study of the two-stage hierarchical model of motion integration where component motions are integrated into a global motion. The Intersection-Of-Constraints (IOC) model, originally presented by Adelson and Movshon, hypothesized that component motions are initially processed separately in V1 before being combined in a weighted sum of the velocity vectors at a later stage, such as in the hMT+ complex. According to this model, the intersection of the two constraint lines orthogonal to each of the component velocity vectors determines the velocity and direction interpretation common to both components, the so-called IOC vector, that predicts the perceived direction of the plaid (Fennema & Thompson, 1979) (Fig. 5b). However, in some cases, the IOC rule cannot predict the perceived motion of the plaid (Fig. 5c). Indeed, it seems that the IOC model only applies to symmetric (or “type I”) plaids for which the IOC velocity vector lies between the velocity vectors of the two component gratings (Ferrera & Wilson, 1990, 1991).

Adaptation-driven models

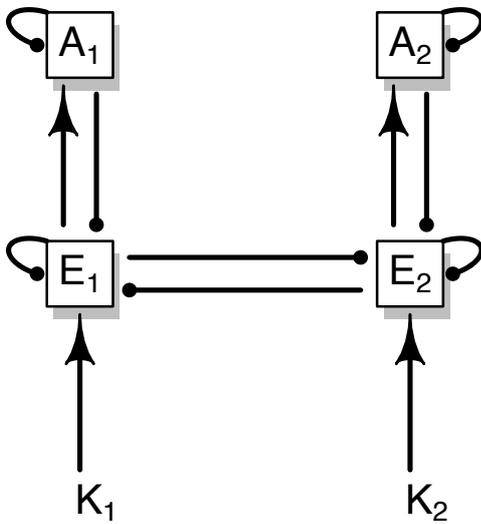
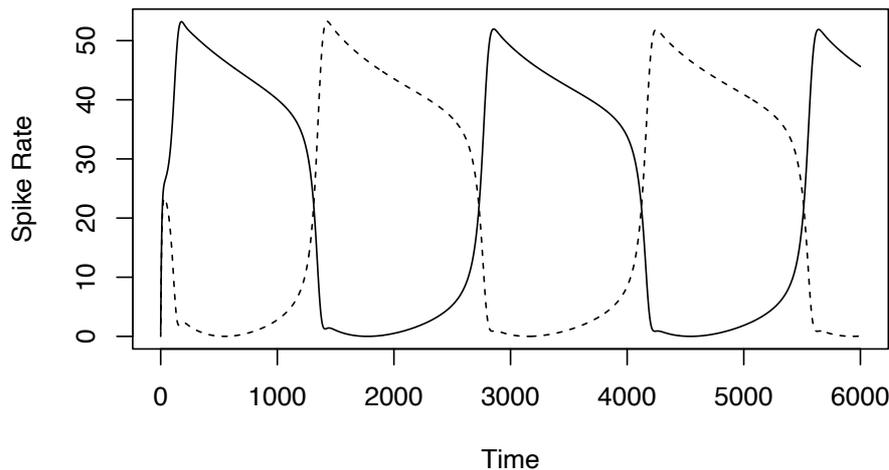


Figure 7: Wilson's model of bistability via adaptation
Schematic representation on the left illustrates:
Excitatory inputs as indicated by arrows;
Inhibitory inputs by black circles
E1 and E2, two excitatory neural units
that mutually inhibit each other,
receive different signal inputs (K1 and K2, respectively),
hypothetically corresponding to distinct perceptual states.
The mutual inhibition results in the more strongly
excited E unit suppressing the other one.
Each E unit excites an adaptation site with
a slow time constant (A1 and A2 respectively)
that will result, through divisive inhibition,
in the eventual decay of response of the active E unit,
so that the other E unit will win out, thus reversing the trend.
The self-inhibitory loops account for the decay of response
to a transitory input at each site in the model.
The bistable alternation of E1 and E2 units
over time is plotted below for a case in which
the time constant of the A units is 30 times greater
than that for the E units.



Evidence against the Adelson and Movshon model has been obtained in studies using asymmetric (or “type II”) plaids (narrow angle, $< 90^\circ$), in which the predicted IOC velocity vector falls outside that of the two components’ (Yo & Wilson, 1992). Experiments with such plaids demonstrate that, during the initial period of the stimulus presentation (≤ 60 ms), the perceived direction is not consistent with the IOC prediction. Instead, within this short interval, the perceived direction that is reported points towards the vector sum (VS) of the two component velocity vectors (Wilson et al., 1992; Wilson & Kim, 1994) before it switches to an approximation of the IOC vector (while maintaining a $\sim 5^\circ$ bias towards one of the components). These studies proposed that the change in the perceived direction reflects a transition between two motion detection systems, namely luminance-based and feature-based, that processed differently motion signals over time. They later extended this model to bistable perception, by presenting perceptual switching as a competition between neural units (i.e., neurons or populations of neurons) encoding distinct percepts and undergoing both reciprocal inhibition and self-adaptation (Fig. 7). In this interpretation of the bistable perception, Wilson et al. suggested that the unit with the higher initial response suppresses the output of the other

unit. In parallel, it was proposed that the activated unit also generates a slow build-up of adaptation that influences the unit's gain (e.g., divisive gain adaptation). At some point, the adaptation becomes sufficient so that the other unit's activity suppresses the adapted unit and wins out, commencing its own build-up of adaptation, and the cycle repeats (Wilson, 1999).

Shapiro et al. (2007) discuss an alternate model from Laing and Chow (2002) that achieves bistability by using subtractive inhibition rather than divisive gain adaptation, but it results in similar dynamics. Interestingly, Weilnhammer et al. (2017) have sought to substitute the notion of accumulation of evidence in place of adaptation and provide arguments why this might describe the dynamics better for a different bistable motion phenomenon.

Noise-driven models

The preceding theories all have in common that they identify adaptation as the main explanatory variable in perceptual alternation, thus defining it as a periodic function of time. To account for the irregularity of state transitions, noise is included in the model. In some models, noise becomes the essential driving force for the switching mechanism (Moreno-Bote et al., 2007; Shapiro et al., 2009). Another theory is that adaptation and noise would have distinct and complementary effects: while adaptive mechanisms would drive the direction of the switch (i.e., perceptual choice), noise would influence the timing of the switch (Huguet et al., 2014).

Another possibility to obtain irregular switching rates is through chaotic dynamics. For example, Strogatz (2018), in his textbook, presents a forced double-well oscillator whose dynamics might account for irregular switching. The model is a mechanical system driven by a sinusoidal input that for certain parameter value switches irregularly between two states. It would be interesting to explore if a neural analogue of such a model could be developed.

A key feature in plaids: the blob

An alternative theory was developed to explain the failure of the IOC model to predict the perceived direction of Type II plaids. It was noted that, because of the narrow angle, such plaids contain elongated high-luminance features referred to as *blobs*, which form at the intersections of the grating components. The blobs are rated as the most salient feature of the plaid for the human eye, and seems to be detected based on their longest edge (1D) and end-points (2D) (i.e., the *interblob* crossing-point with lower luminance) (Fig. 8). Note that blobs are present at the intersections of symmetric gratings, also, but they are generally less salient. Such observations led to the proposition that local motion detectors, existing as early as area V1, could respond to the motion direction of both the blob (orthogonal to the orientation of the longest edge) and the end-point features (Wilson et al., 1992; Dimova & Denham, 2010). Far from the traditional 2-stage feedforward model of motion integration (IOC model) described as a weighted sum of first-order motion components, this feature-based model explains motion integration in terms of the blob/end-point features. According to this model, the plaid is defined as a product of (anti-phase) component gratings rather than a sum. It posits that the early misperception is dominated by 1D motion signals (along the high-luminance long edge of the blob), before 1D and 2D (end-point) signals are combined in a recursive Bayesian integration mediated through reciprocal interactions between V1 and hMT+. Such recurrent interactions, which are known to be strongly involved in perceptual awareness of motion (Sterzer et al.,

2006), would lead to a progressive shift in the dominant form of motion perception, from 1D to 2D motion.

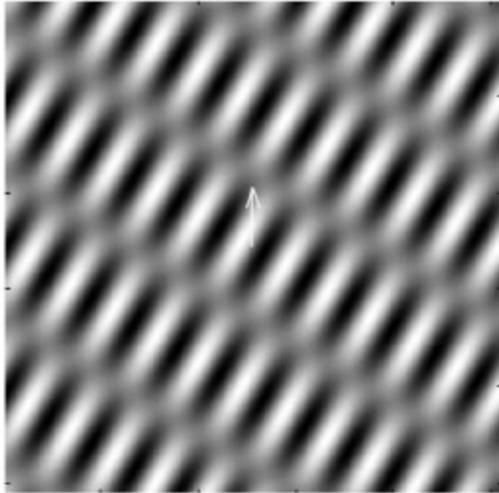


Figure 8: A plaid containing blobs as illustrated by Wilson et al., 1992

4. Summary

The goal of this first chapter was to introduce some of the key findings that reveal a hierarchical structure of visual cortex in cats and primates. Ascending and descending projections appear to follow general statistical rules both in terms of the distribution across cortical layers and inter-areal patterns of connectivity. Oscillatory coherence between areas has been shown to follow the anatomical laminar pattern, thereby suggesting a matching functional hierarchy. The feedforward visual hierarchy is strongly supported by the role of receptive fields in automatically generating holistic representations based on local processing of sensory signals. This contrasts with feedback which is much more difficult to study. On the other hand, if we consider global-to-local processing (e.g., we consider the whole face before examining its inner components, Peters et al., 2018), Hochstein and Ahissar in 2002 literally reversed the hierarchy by arguing that holistic representation can precede the examination of local details of interest (i.e., seeing the “forest before the trees”) through active/conscious perception. Moreover, this reverse hierarchy theory is consistent with predictive processing theory, showing that internal global representations (i.e., predictions) of the visual environment are refined through recursive local active sensing.

In my thesis project we studied a well-known hierarchical model of visual perception: motion integration. We implemented a paradigm of bistable perception that enabled investigating changes in perception independently of visual input (i.e., the stimulus parameters remained fixed). Therefore, we hypothesize that, due to the ambiguity inherent to the stimulus, the spontaneous perceptual switching that occurs will reflect a competition between high-level and low-level representations of motion.

The first step of my thesis (described in Chapter 2) will be designing a bistable moving plaid that alternates perceptual state at approximately regular intervals, with intervals long enough to permit studying brain states through fMRI. This justifies the review of different motion detection systems characterized in literature, as well as the physical properties of bistable plaids. In this part, I will present results and analyses of a psychophysical experiment that we

designed in order to optimize the parameters allowing bistable alternation between moving components and pattern perceptual states.

Then, using this bistable plaid, we investigated the neural correlates of perceptual decisions with an event-related fMRI paradigm, by tracking motion integration signals in areas V1, V2 and hMT+, as a function of perceptual state (as reported by observers). This second project, developed in Chapter 3, reveals evidence of a higher-order motion representation at the lowest level of visual cortical hierarchy, V1.

Finally, we explored the influence of eye movements on the neural correlates of bistable perception. The question is whether or not the previous results were not biased by eye movements. Therefore, the Chapter 4 of my thesis analysed the variability of eye in the two perceptual states, first comparing eye movements made prior to perceptual switches, and also during the reported percept.

CHAPTER 2: **The Dynamics of Bistable Perception**

In the introduction, I described several computational models of neural competition that aimed to illustrate the phenomenon of bistable perception. The aim of this thesis is to detect whether neural correlates associated with bistable motion perception are detectable in early visual areas. Given the known physiology of these areas, a positive finding would support the presence of feedback processes associated with this phenomenon. Thus, it will be necessary to define stimulus conditions that generate bistable motion perception in a reliable and reproducible fashion. In this chapter, I will describe the psychophysical experiments that were performed to explore parameters that produced the bistable phenomenon. What we seek are stimuli whose appearance over an extended viewing period alternates between two different perceptual states with an approximate regular period between transitions. It is also important that the alternation rate is not too rapid, i.e., that each of the states remains stable for a sufficient time to enable measurement of BOLD responses associated with each state in the subsequent experiments, but it should be short enough to measure several perceptual transitions over the viewing period. For example, with a one-minute viewing period, 12 to 20 alternations would indicate that each perceptual state was stable for between 3 and 5 seconds, on average.

1. Psychophysics: How to design a bistable plaid?

Inducing a bistable perception requires designing a stimulus that will trigger perceptual switching between two equiprobable interpretations of the same motion information. The example of bistable perception that will be studied here is constructed from superimposed one-dimensional component gratings of different orientations, moving in different directions. Prolonged viewing of these stimuli leads to spontaneous switches between two different perceptual states: perceiving the component gratings moving in different directions, sliding transparently over each other, or an integrated plaid pattern moving in a single direction different from each of the components. Initial experiments were run to find stimulus parameters for which each perceptual state lasted more than three seconds and the average amount of time spent in each state was about equal. Several behavioral studies are in general agreement about critical parameters for facilitating bistable plaid dynamics. First of all, designing asymmetric plaids by creating a situation where component gratings have different physical attributes (e.g., spatial frequency, duty-cycle, speed, orientation difference of the gratings or angle, contrast etc.) generally enhances the transparency effect, therefore allowing perceptual switches towards the component percept (Hupé et al., 2019; Huguet et al., 2014). In order to find a balance between the dominance of plaid and component percepts, the transparency effect can be manipulated and counterbalanced by setting the contrast between grating's intersections and other parts of the plaid in an in-between situation between transparent and non-transparent regimes. To implement this, we used two of the most influential variables to design a set of plaids varying systematically with respect to the values of two parameters: the orientation difference or angle between the component gratings and their difference in contrast (Von Grünau & Dubé, 1992 ; Stoner & Albright, 1996 ; Hupé & Rubin, 2003 ; 2005 ; Klink et al., 2008; Huguet et al., 2014). The most common experimental method to evaluate this is to measure the distribution of the durations of the two perceptual states as a function of angle and contrast levels.

1.1 General properties of bistable plaids

Brief introduction

Behaviorally, bistable perception is defined as a spontaneous alternation of perception between two equiprobable perceptual states (although no physical change actually occurs). Moreover, in order to produce such ambiguity, the stimulus has to meet specific criteria. In the example of bistable plaids, such a situation can be arranged by setting specific physical parameters (described below) and by presenting the stimulus for long periods of time to the observer. Various psychophysical studies have identified the principal parameters influencing bistable perceptual dynamics, in particular altering the angle between the component gratings, the relative transparency and speed.

Plaids with transparent grating components can induce up to three percepts, hence they are often referred to as tristable, alternating between the pattern percept (i.e., a coherent motion of the plaid), and the transparent component motions with either one or the other grating in the foreground (Hupé & Pressnitzer, 2012). Hupé, Signorelli and Alais (2019) were the first to modulate separately the strength of these percepts. They interpreted their results by proposing that each of these three percepts is encoded by distinct neural populations that are mutually inhibiting each other.

Effect of depth-ordering

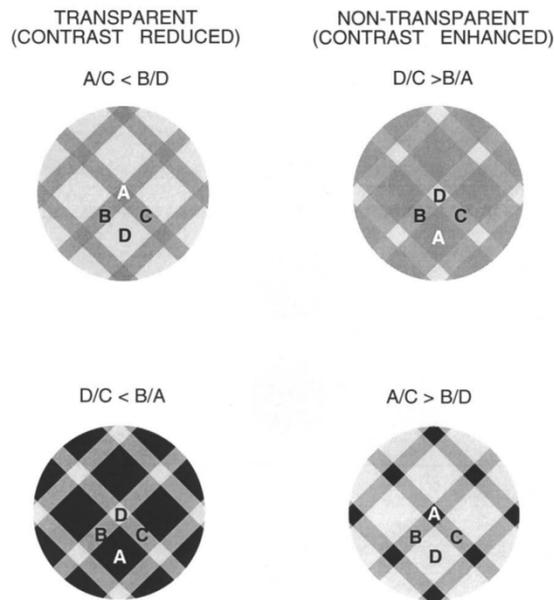
Within this framework, we chose to study plaid bistable motion perception, a visual illusion used to study motion integration phenomenon, which actually also concerns monocular depth perception. Therefore, we limited the perceptual alternation to two options (bistable perception), either experiencing the motion of the plaid or perceiving one grating (always the same) in front of the other. For this purpose, we were able to replicate Hupé et al.'s method by constraining the depth-order (i.e., transparency relation) to one unambiguous component state (foreground grating perceived as occluding background grating).

I will briefly introduce a few properties derived from the physics of transparency (Metelli, 1974; Beck et al., 1984) that indicated which cues drive depth perception when presenting simple moving structures. The underlying mechanisms of depth-ordering involve very general effects of transparency that were extremely useful in determining how to design a bistable plaid (the main objective of Chapter 2 psychophysics).

Three main depth cues are characterized in asymmetric plaids (i.e., plaids made of two grating components with distinct physical properties). The most influential one is the ratio between the spatial frequencies of the components (i.e., wavelength): the grating with a shorter wavelength (higher frequency) tends to be perceived at the back. To a lesser extent, duty cycle (i.e., light/dark cycle width ratio) and speed effects have also been described, showing a tendency to perceive the grating with the smallest duty cycle or the highest speed as behind the other one (Moreno-Bote et al., 2008). However, I will not enter into the detail of these effects as we did not exploit parameters that favored the component percept at the expense of the pattern percept.

Instead, we exploited the rules governing luminance ratios between the different regions of the plaid, an important factor also shown to influence depth perception (Stoner & Albright, 1996, Stoner et al., 1990), and thus plaid's motion integration. What mattered was the relative

contrast between the intersections and the rest of the plaid more than the absolute values of luminance itself.



For example, if the objective is to avoid the transparency effect (which favors the perception of two sliding components), then it is required to display the intersections in such a manner that they pop out from the rest of the plaid by enhancing the contrast of the intersections (see fig. 9, non-transparent regime).

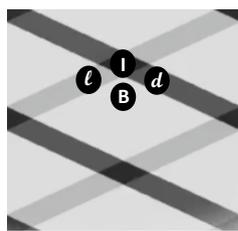
The transparency effect is governed by the depth-ordering between the two component gratings. A plaid is considered “transparent” if the intersection contrasts relative to each grating line is reduced (see fig. 9, transparent regime).

Figure 9: Schematic depiction of the transparency effect in plaid's motion perception described by Stoner & Albright (1990; 1992ab; 1996, 1998)

Here are four plaids differing in the contrast ratio between the intersection (A), both gratings' line (B and C) and the background (D) regions. Each case illustrates a different relation of luminance contrast between regions A, B, C and D, in a way that it alters surface segmentation and depth perception.

Behaviorally, a reduced contrast (i.e., the most salient contrast in the stimulus corresponds to the one between gratings' thin bars and the background) tends to induce a transparent perception of motion (i.e., the components), whereas an enhanced contrast of the intersections (in relation to the gratings' bars) makes the plaid's motion perception dominant.

An extrapolation to the previous rule can be made when dealing with asymmetric plaids, by setting one grating in enhanced contrast relative to the intersection (perceived at the back), while the other is in reduced contrast (perceived on top). Figure 10 illustrates the luminance of the intersection region (I), the background (B), and gratings' thin bars (ℓ and d). Stoner and Albright demonstrated that the more similar the grating' stripe luminance (d) is to I and the more it differs from B, the more likely component motion will be perceived (“reduced contrast”). In contrast, the greater the difference between gratings' thin bars luminance (ℓ) and I and the more similar to B, the more it favors pattern (integrated) motion perception (“enhanced contrast”). Therefore, by setting one grating to reduced, the other to enhanced contrast, we hypothesized that it would maximize the ambiguity in the motion percept.



Coherence (enhanced contrast):
 $l/\ell > l/B$
 Transparency (reduced contrast):
 $l/d < d/B$

Figure 10: Transparency versus Coherence, a matter of depth-order
 Component motion perception is facilitated by transparency between gratings. A way to implement this is to vary the contrast of intersection region (I) relative to the rest of the plaid (l, d, B). Reducing this contrast will more probably elicit the component percept, whereas enhancing it has the opposite effect (pattern).
 For this reason, we created a hybrid plaid with one grating in reduced, the other in enhanced contrast regime (Stoner & Albright, 1990; 1992ab; 1996; 1998).

Effect of angle

The angle α between component gratings has been shown to have a significant influence on the tendency to perceive coherent versus transparent motion in many research studies (Adelson & Movshon, 1982 ; Kim & Wilson, 1993 ; Hupé & Rubin, 2003). All of them agree that angles (α) closest to 90° facilitate the perception of a coherent motion. And conversely, the more the angle differs from a square angle, the more transparent motion will appear.

Effect of speed

The effect of speed, although significant, appears less determinate than the contrast and angle parameters. Many studies have shown that increasing speed facilitates motion transparency (as it provides a higher level of motion contrast between the two component directions), however this only applies to a very limited range of speeds (high speed) and depends on the angle α (Farid & Simoncelli, 1995 Smith, 1992; von Grunau & Dubé, 1993; Hupé et al., 2019).

1.2 Motivations

The main objective of this behavioral experiment is to determine the stimulus parameters that optimize the probability of switching percepts, i.e., perceiving the plaid's motion as “coherent” (i.e., pattern), or “transparent” (i.e., components) 50 % of the time. As described in the introduction, it was shown that increasing the angle between gratings ($>90^\circ$) strengthens the component percept, whereas the pattern percept dominates when the angle approaches 90° .

Secondly, enhancing or reducing the contrast of the intersections relative to grating' stripes has been shown to affect motion integration (Fig. 10). Moreover, another argument for setting

different contrasts between gratings is that it enabled maintaining the perceived depth-order constant for the component percept (with stripes d on top of ℓ), hence constraining the range of perceptual states to a bistable perception (component/pattern) instead of a tristable plaid, (Hupé & Pressnitzer, 2012).

1.3 Methodology

Participants

Six healthy volunteers (3 women, 3 men, aged 18 to 30 years) with normal or corrected-to-normal vision were included in the study. They all gave informed consent and performed 5 sessions. However, one of them (the experimenter) was considered an outlier and excluded from the group-level analyses. All other participants were naive to the experiment.

Stimuli

Specifically for this experiment, we created a set of 25 plaids varying in both angle (α) and contrast between grating components (Δc) (5x5 factorial design matrix, with 5 levels of α crossed with 5 levels of Δc) (Fig. 11). α levels ranged from 90° (presumed to induce strong pattern percept) to 140° (strong component percept). Similarly, Δc varied from 0% (identical contrast between gratings) to 60% (i.e., one grating's contrast is enhanced (80%), while the other is reduced (20%)).

Except for the stimulus with identical contrast ($\Delta c = 0\%$), one grating had its thin bars always more similar to the intersection's luminance than the other. Setting this difference of contrast is known to influence perceived depth-order (i.e., one component is perceived as sliding on top of the other).

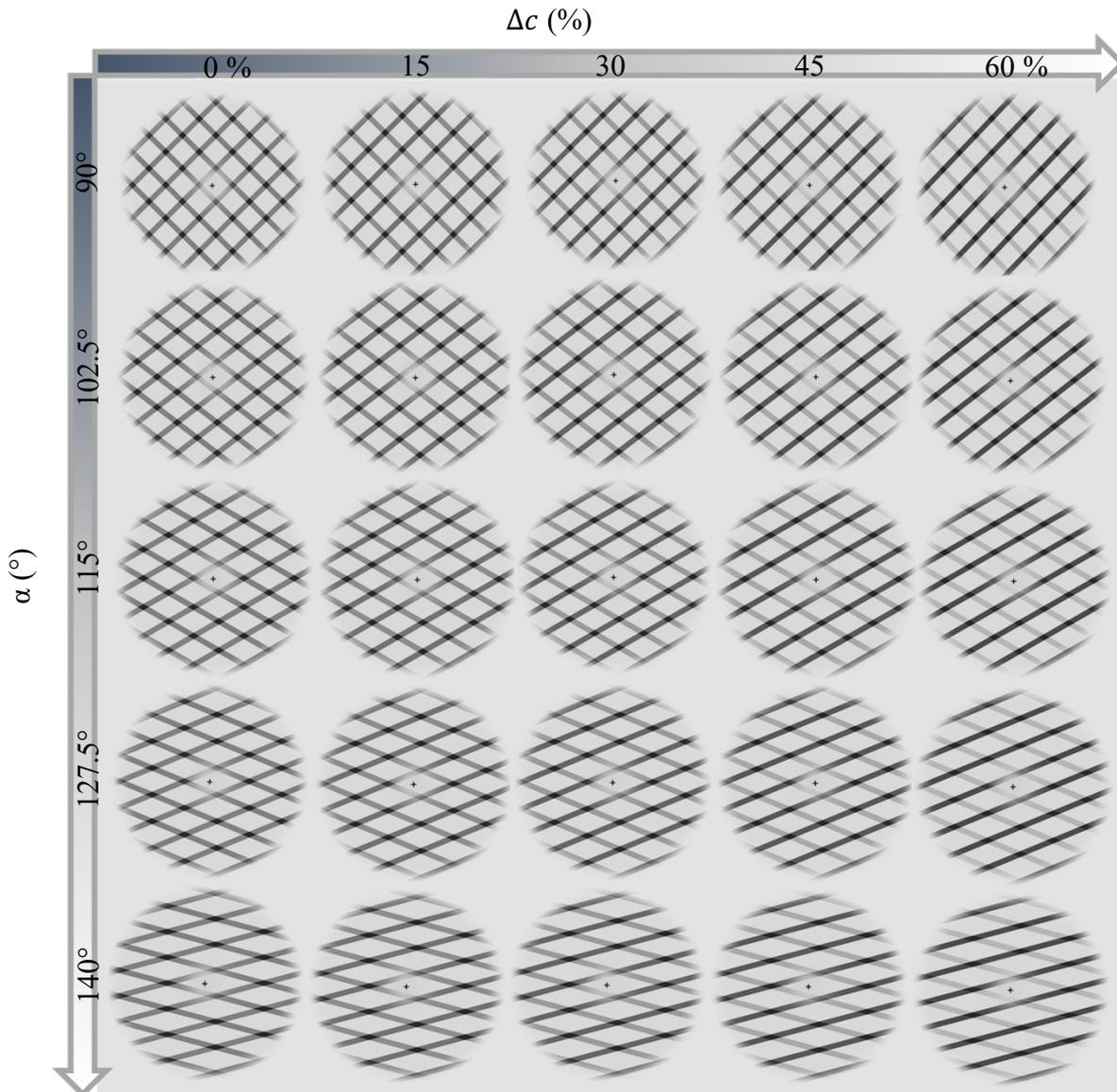


Figure 11: Plaids differing in angle and contrast levels

The 5x5 design matrix was used to test for the effect of angle and contrast between grating components. In this experiment, each plaid was shown for a short duration (400 ms) and repeated 50 times.

Apart from the two aforementioned variables, all plaids shared similar structure: stimuli were presented in a circular annular field whose outer border subtended 13 degrees of visual angle and inner border 1.5. The outer border was masked with a raised cosine and the inner with a Gaussian to smooth the edges. A black fixation cross subtending 0.5 deg appeared in the center. The background was grey equal to the space average luminance of the grating with CIE coordinates ($xyY = (0.33, 0.34, 100 \text{ cd/m}^2)$). The stimulus consisted of two non-colinear overlapping square-wave gratings, each moving orthogonally to its orientation. Both component gratings had spatial frequency of 0.6 cycle/deg, with unequal duty cycle (22.2 %), moving at a speed of 1.7°/s. Moreover, grating components were spatially arranged symmetrical to the horizontal axis (0°).

Each trial consisted of a short presentation of the moving plaid (400 ms) interleaved with a 600-1200 ms response period between stimuli. 5 sessions were performed, resulting in 50 repetitions of each combination of α and Δc .

Observers were positioned 57 cm from the screen in a darkened room equipped with a chin rest). While fixating a central target cross, they were presented a plaid with its characteristics randomly chosen from the stimulus set and responded as to whether motion was perceived in one direction (pattern) or two (component). If we measure the proportion of trials on which they respond “one” (pattern), this is equivalent to a Yes/No task. Observers were instructed to maintain fixation during the whole session.

Contribution of angle and contrast to the perceived motion

A first step was to visualize the contribution of α and Δc variables in each observer’s motion perception (observer IDs: P-L-A-I-D). To implement this, we measured the proportion of pattern percept over component ($P_{P/C}$) as a function of α and Δc (e.g., $P_{P/C} \rightarrow 0$ was indicative of 0 % of pattern / 100 % component percept and inversely). In other words, one can consider this as an indirect measure of the % of motion integration as a function of the stimulus physical properties.

Individual 3D perspective plots illustrate the two explanatory variables in the same referential, and their 2D plane projections showed that both angle and contrast influence perceived motion. Indeed, for each observer, Fig.12 (column 2) shows the effect of Δc for each level of α_i . Conversely, Fig. 12 (column 3) illustrates the effect of α across Δc_i levels. The steeper the curve is, the higher the effect of the continuous variable. A flat line (coefficient 0) is indicative of a null effect of the continuous variable, i.e., that the influence of one variable is independent of the other.

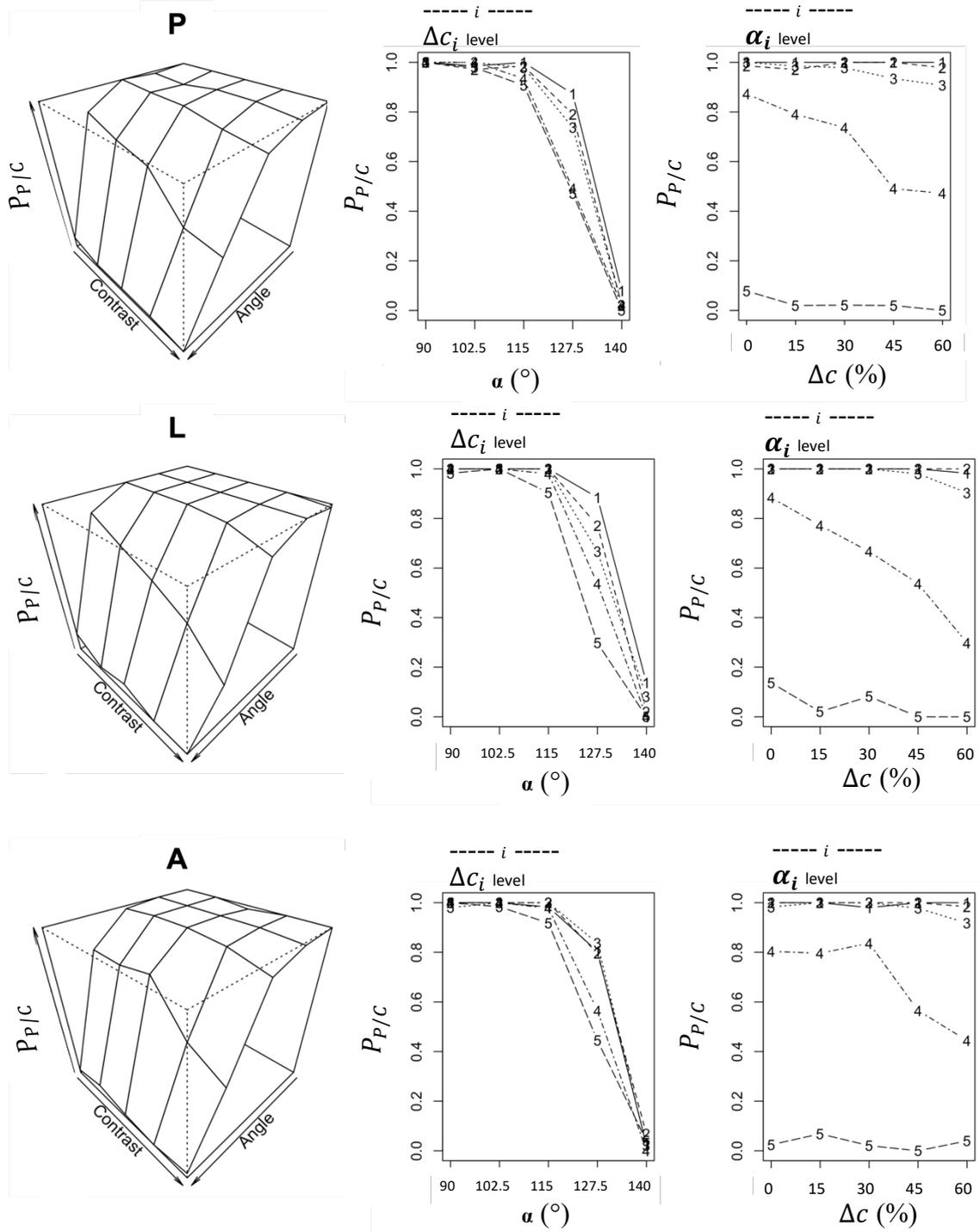


Figure 12: Contribution of angle and contrast to perceived motion

On the left (column 1), individual 3D perspective plots representing the proportion of perceived pattern (vs component) motion as a function of contrast and angle settings as reported by $N=5$ observers (P-L-A-I-D). Δc corresponds to the difference of contrast of gratings each in relation to the intersection region. α corresponds to the angular difference between the two gratings. The same information is shown in 2D plane, representing the effect of α gradient across Δc_i levels, and the effect of Δc gradient across α_i levels.

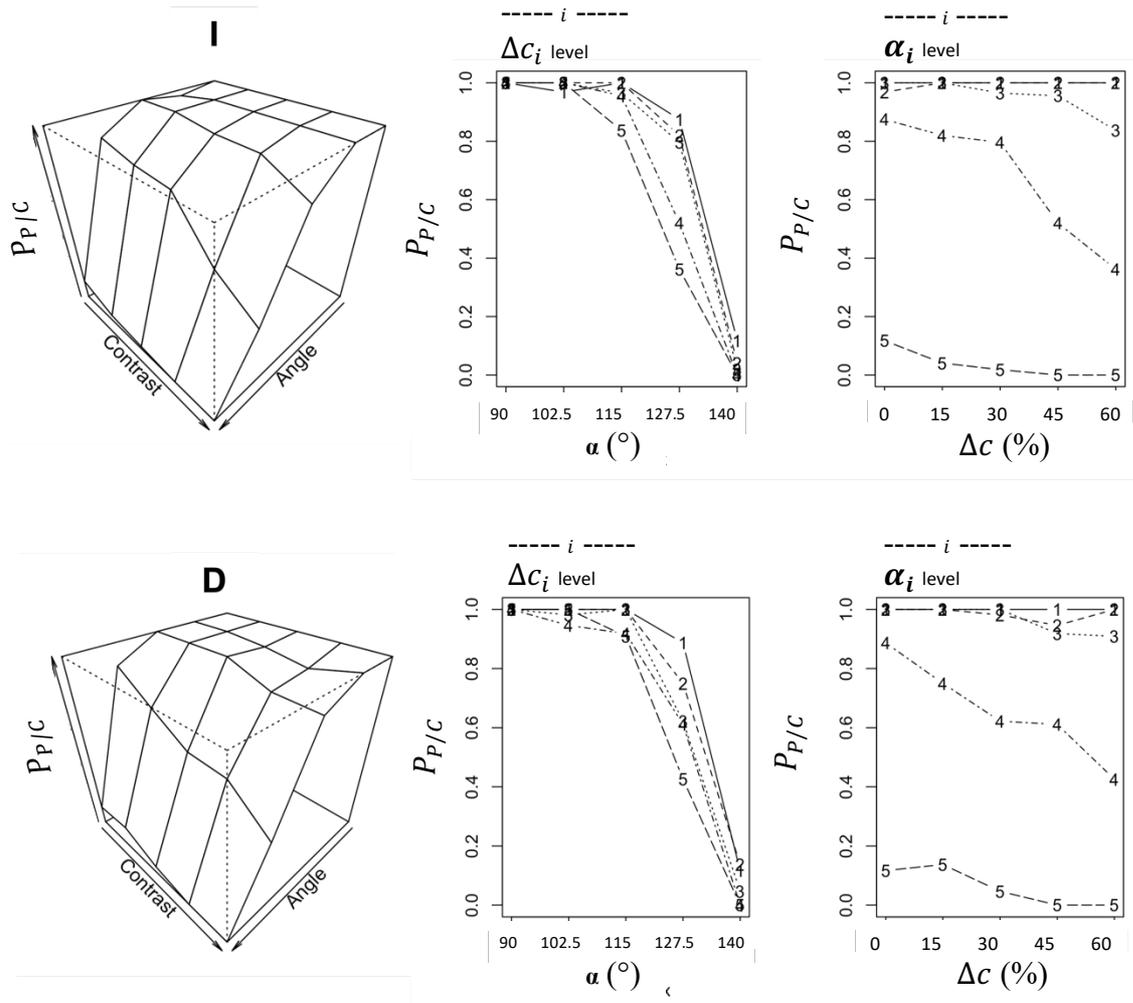


Figure 12bis: Contribution of angle and contrast to perceived motion

There is remarkable agreement between observers, overall. Both variables influenced perceived motion, with the major effect attributed to α (independently of Δc_i level). Indeed, the α effect on perceived motion showed a steep variation and a full coverage of $P_{P/C}$ amplitude. In comparison, the motion percept was relatively independent of Δc except for level 4 of α .

In order to generalize the relation between these stimulus physical properties and the perceived form of motion, we investigated α and Δc at the group-level.

The objective was to predict the probability associated with response variable (“observer reported pattern (=1) or component (=0)”) as a linear model of α and Δc predictor variables, hypothesized to influence the way stimulus was perceived. A suitable model for this type of dataset is a binomial Generalized Linear Model otherwise known as logistic regression. Note that the Generalized Linear Model, often abbreviated by GLM in the literature, is different from the General Linear Model, also abbreviated by GLM, used below to analyse fMRI results. To avoid confusion, we will use lower case glm for the Generalized and uppercase GLM for the General Linear models, respectively. This makes sense as the function in the R programming language used to fit this model is in lowercase.

1.4 Introduction to binomial Generalized Linear Model (glm)

Definition

Each trial can be considered as an independent Bernoulli event whose outcome refers to observer's response ($Resp = [\text{pattern}=1, \text{component}=0]$). This response variable is a special case of the binomial distribution with $n = 1$:

$Resp \sim B(n_i, P_{P/C})$, where n_i represents the number of trials ($n_i = 1$, here)

$P_{P/C}$, the proportion of $Resp = 1$,

with variance $var(P_{P/C}) = \frac{1}{n_i} P_{P/C}(1 - P_{P/C})$,

that can be simplified to $V(\hat{\mu}_{P/C}) = \hat{\mu}_{P/C}(1 - \hat{\mu}_{P/C})$,

with $\hat{\mu}_{P/C}$ estimating the probability that $Resp = 1$

Next, predicted values are transformed on a $0 \rightarrow +\text{Inf}$ scale representing the odds ratio, that is the ratio between the probability of $Resp = 1$ and $Resp = 0$.

$$odds_{P/C} = \frac{\hat{\mu}_{P/C}}{1 - \hat{\mu}_{P/C}}$$

Finally, the *logit* function is given by the natural logarithm (base e) of the odds ratio:

$\eta_{P/C} = \text{logit}(\hat{\mu}_{P/C}) = \log(odds_{P/C})$, with $\eta_{P/C}$ predictions now ranging from $-\text{Inf} \rightarrow +\text{Inf}$. The $\eta_{P/C}$ is a linear model (or predictor) of the explanatory variables α and Δc that is related to the expected probability of choosing the pattern stimulus through the logit transform. The contributions of both variables are estimated by maximum likelihood using the glm function in the OpenSource software R. The expected probabilities are obtained by inverting the logit function to obtain:

$$\hat{\mu}_{P/C} = \frac{\exp(\eta_{P/C})}{1 + \exp(\eta_{P/C})}$$

The significance of the contributions of the explanatory variables is evaluated by comparing nested models using likelihood ratio tests.

How to plot the binomial glm?

As previously described, the estimated probability $\hat{\mu}_{P/C}$ is expressed as a linear combination of predictors by means of the *logit*, a sigmoid link function (see Fig. 13 for example).

This is an example of a psychometric function, a family of psychometric models applied in detection and discrimination tasks with choice paradigms. The data points are obtained by aggregating the responses within specific combinations of angle and contrast and averaging the response, which yields the proportion of "pattern" responses.

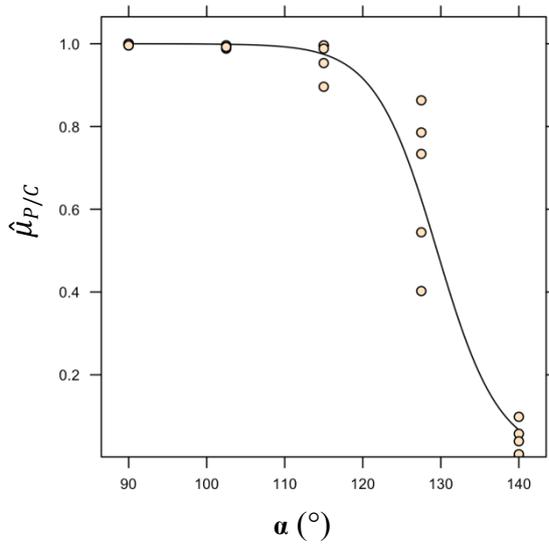


Figure 13: Pattern versus Component discrimination as a function of angle
Psychometric functions of the probability of pattern (1) versus component (0) responses as predicted by glm linear model $\hat{\mu}_{P/C} \sim \alpha + \varepsilon$
N= 5 observers

Comparing glm's goodness of fit and complexity

Because individual profiles showed that $\hat{\mu}_{P/C}$ mainly varies as a function of α , and that this contribution seemed to be independent of Δc , we first chose to focus on the main effect α . However, to keep sense of Δc effect, we defined a binomial GLM $\hat{\mu}_{P/C} \sim \alpha + \varepsilon$ independently for each Δc_i level (Fig. 14).

Fig. 14 represents model predictions $\hat{\mu}_{P/C}$ estimated as a function of α and compared across Δc_i levels. The steeper the curve of the psychometric function is, the higher the absolute value of α slope coefficient, and thus the sharper the transition between the two perceptual states. To obtain equiprobable transition probabilities, the contrast level should be chosen where the psychometric function gives a value of 50%.

Based on the model predictions, it appears that the absolute estimated coefficient for α (continuous variable) increases with contrast level, Δc_i (ordinal variable). To test this hypothesis (H_1) against the null hypothesis of “no significant effect of contrast level on the predicted motion perception”, we defined and compared the following nested models.

The simplest model (1) expresses $\hat{\mu}_{P/C}$ as a function of continuous variable α exclusively (H_1). Model (2) adds a linear effect of Δc_i (H_2 : “additive effect of contrast level”). Finally, the most complex model (3) tests for potential interaction effects between α and Δc_i level (H_3 : α slope coefficients depends on contrast level).

- (1) $\hat{\mu}_{P/C} \sim \alpha + \varepsilon$ (independent model: only α)
- (2) $\hat{\mu}_{P/C} \sim \alpha + \Delta c_i + \varepsilon$ (additive model)
- (3) $\hat{\mu}_{P/C} \sim \alpha + \Delta c_i + (\alpha : \Delta c_i) + \varepsilon$ (interaction)

The best model was selected after a tradeoff between goodness of fit and model complexity with an Analysis of Deviance. The additive model (2) came out being highly significant (Analysis of Deviance, $p < 2.0 \times 10^{-16}$) and showed the smallest Akaike information criterion (AIC, indicator of prediction error). It indicated that $\eta_{P/C}$ linear predictor was best modeled as a sum of α and Δc_i effects (rejecting H_1), and that there was no significant interaction effect between the two explanatory variables (rejecting H_3).

2. Group-level statistics for the additive model

Each term of the model reached high significance as reported in the table of coefficients below (table 1). The estimated coefficient for the intercept corresponds to the condition/plaid with $\alpha = 90^\circ$ and $\Delta c = 0\%$, that theoretically corresponds to the strongest pattern state.

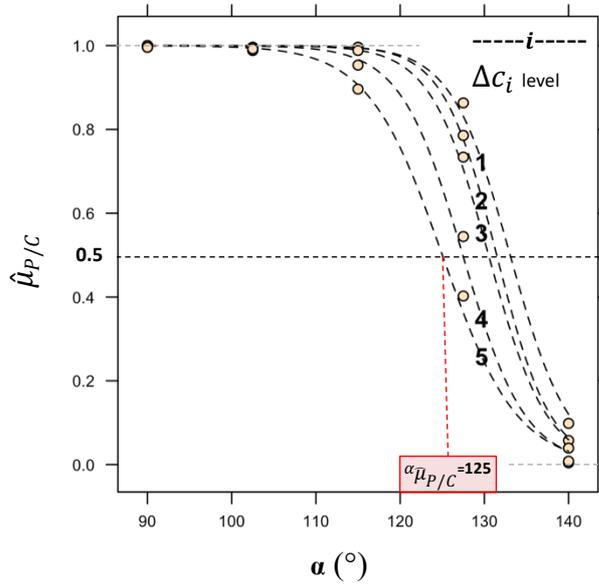


Figure 14: Pattern versus Component discrimination as a function of angle across contrast levels
Psychometric functions of the proportion of pattern (1) versus component (0) percepts as predicted by binomial glm (2)
 $N = 5$ observers

Coefficients :

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	36.325384	1.108037	32.784	< 2e-16 ***
Δc_2 level	-0.401343	0.170901	-2.348	0.0189 *
Δc_3 level	-0.694107	0.169994	-4.083	4.44e-05 ***
Δc_4 level	-1.529787	0.170158	-8.990	< 2e-16 ***
Δc_5 level	-2.197581	0.177241	-12.399	< 2e-16 ***
α	-0.272929	0.008314	-32.826	< 2e-16 ***

Table 1 : Estimated coefficients of angle and contrast levels

Summary of binomial glm (2)

Each variable's z-value is tested against the intercept ("strongest pattern state").

Negative values indicate a shift towards the component perceptual state.

All effects of α and Δc_i were highly significant (***, $p < 0.001$)

except for lowest levels of Δc_i (*, $p < 0.05$).

In comparison to this presumed to be "strongest pattern state", coefficients were estimated for continuous variable α , confirming that it was the most influent factor in explaining $\hat{\mu}_{P/C}$ variation ($\hat{\mu}_{P/C}$ decreased as α increased, z-value = -32.8, $p < 2.0 \times 10^{-16}$).

Moreover, table 1 shows that the pattern-to-component transition reached the largest coefficient (steeper slope) for high contrast differences (for $\Delta c_5 = 60\%$, z-value = -12.4, $p < 2.0 \times 10^{-16}$), also confirming our previous observation (Fig. 12).

On top of showing a steeper pattern-to-component discrimination slope, Δc_5 also ensured a more complete coverage of $\hat{\mu}_{P/C}$ amplitude (from 1: strongly pattern-like, to 0: strongly component-like).

Finally, after determining the optimal level of contrast difference ($\Delta c_5 = 60\%$), we reported the associated value of the angle parameter as $\alpha_{\hat{\mu}_{P/C}=50\%} = 125^\circ$, where $\hat{\mu}_{P/C} = 50\%$ represents the

intermediate state presumed to equiprobably lead to either of the two concurrent perceptual states (pattern/component).

3. Conclusion

This experiment motivated our choice for optimal angle and contrast settings that would perceptually maximize plaid's bistable motion. Moreover, this alternation of perception (ideally between the two principal perceptual states) had to be the most direct, by limiting the number of intermediate states that would constitute a factor of noise in our fMRI study of perceptual decision.

Varying the angle between gratings satisfied this criterion by showing strong pattern-to-component dependence and a full coverage of the two strongest perceptual states. The effect was weaker and more variable with the contrast parameter.

However, when comparing the contribution of angle across the different contrast levels, we found an additive effect of contrast level that was specific to the highest level ($\Delta c = 60\%$, consisting of a grating in "reduced" contrast (20%), and the other enhanced (80%)).

With $\Delta c = 60\%$ and $\hat{\mu}_{p/c} = 50\%$ (bistable stimulus), the model predicted a value of $\alpha_{\hat{\mu}_{p/c}=50\%} = 125^\circ$. Altogether, these results were in accordance with previous psychophysical studies investigating prolonged exposure to ambiguously moving plaids (for a review, see Hupé and Rubin, 2003).

Therefore, we designed our bistable plaid ($\Delta c = 60\%$ and $\alpha = \pm 62.5^\circ$ relative to the horizontal axis) (Fig. 10) with the main property of inducing a well-balanced competition between pattern and component perceptual states (Fig. 14). This bistable plaid was the main stimulus used in the fMRI experiment presented in Chapter 3.

CHAPTER 3: **Neural Correlates of Perceptual Decisions**

In this chapter, the neuroimaging results and analyses of the main experiment are presented. In order to investigate the neural correlates of perceptual state transitions in a bistable motion phenomenon, fMRI recordings were obtained from direction-selective subdomains in visual cortical areas V1, V2 and the hMT+ complex while participants continuously viewed a bistable moving plaid stimulus and reported on changes in perceptual state. Since the physical parameters of the stimulus remain unchanged during viewing, the perceptual transitions are unlikely to be mediated by bottom-up signals, and it is hypothesized that they will be related to top-down activity. The results and analysis of the simultaneously recorded eye movement activity will be presented in the subsequent chapter. The current chapter begins with a brief historic review of fMRI recording, its physical bases and the current understanding of its relation to neural activity. This is followed by a brief discussion of the General Linear Model (GLM) used to estimate and test hypotheses relating imaging signals to stimulus conditions. Section 3 describes the stimuli and experimental paradigms for the main experiment and for localizing direction-selective subdomains. This is followed by a section detailing the analysis pipeline. Sections 5 and 6 describe the results and conclusions, respectively.

1. Introduction to functional magnetic neuroimaging

Functional Magnetic Resonance Imaging (fMRI) is a relatively recent technique, developed in the 1990s (credited to Seiji Ogawa and Ken Kwong), to measure brain activity indirectly and non-invasively through the Blood Oxygen Level Dependent (BOLD) signal. The underlying concept is quite intuitive: the more oxygenated blood flow increases/decreases in a specific region of the brain (relative to its baseline activity and to neighboring areas), the more it reflects an increased/decreased need for energy, and thus a local change of neural activity. This general idea was first suggested by an Italian scientist, Angelo Mosso, whose experiment was then published in William James' book *The Principles of Psychology*, in 1890.

‘The subject to be observed lay on a delicately balanced table which could tip downwards either at the head or the foot if the weight of either end was increased. The moment emotional or intellectual activity began in the subject, down went the balance at the head-end, in consequence of the redistribution of blood in his system.’

Although the method sounds very experimental and questionable as it aims to measure variation of blood flow in general and not oxygenation level specifically, the relation made between brain activity and the redistribution of blood, more particularly during intellectual or emotional tasks, was very innovative and constituted a turning point for non-invasive neuroimaging. At the time a common belief was that no local variation of blood flow could occur in the brain tissue due to the fact that it is embedded and constrained by the skull.



Figure 15: The Sherrington Starling Kymograph

Charles S. Roy and Charles S. Sherrington (1890) were the first to show evidence of a coupling between energy metabolism and blood flow in the brain. They conducted a series of experiments on anesthetized dogs by placing a monitoring device, the “Kymograph”, on the cortical surface, measuring blood volume variable (Fig. 15). They were able to measure significant fluctuations of blood flow, however they could not determine the cause of this physiological response.

It was only later, in 1948, that Seymour Kety and Carl Schmidt identified the brain self-regulation of oxygenated blood level as a function of its regional energy cost. Notably, they postulated that because neurons used more oxygen, surrounding blood vessels received chemical signals to vary their diameter, thus regulating cerebral blood flow (CBF). This breakthrough made fMRI an attractive tool to investigate brain function in humans.

1.1 Biology

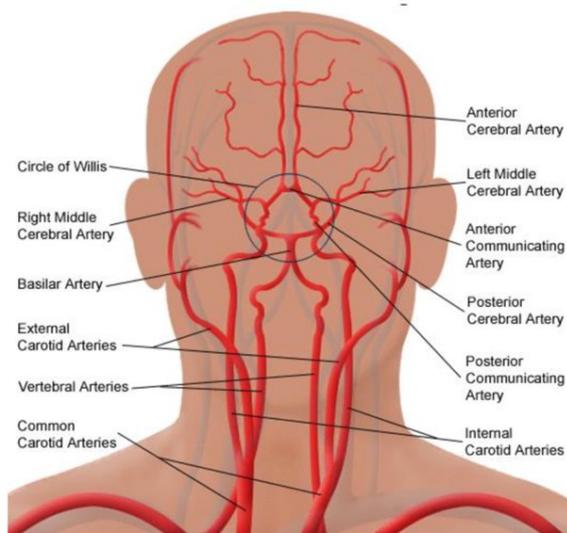


Figure 16a: Organisation of the human cerebral blood circulation (by Yohannes Mamo, 2015)

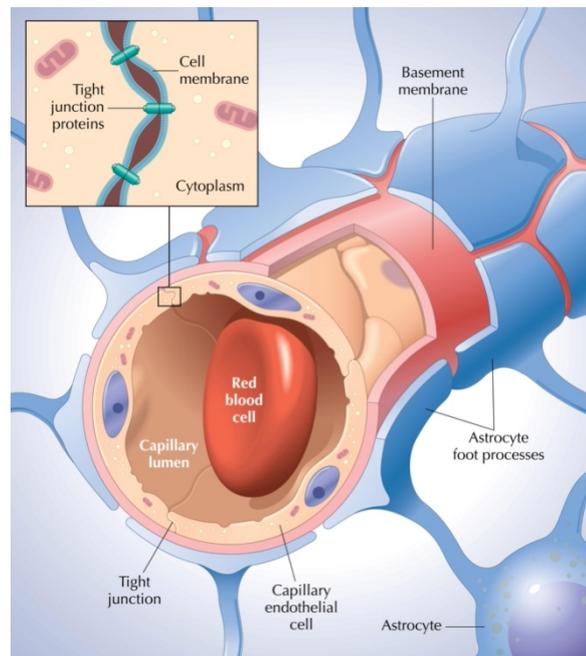


Fig. 16b: Illustration of the Blood-Brain barrier (by James Perkins)

In physiological conditions, oxygenated blood that flows to the brain (~20%) is remarkably constant, mainly due to the prominent contribution of large arteries (carotid + vertebral) to vascular resistance (that regulates blood pressure, i.e., blood vessel length/diameter and blood viscosity). However, brain tissue requires a tight regulation of blood flow adapted to dynamic and local increased demand for oxygen. Such regulation is mediated through vasodilation of

distal/proximal arterial segments. This arterial system leads to a dense and complex capillary system where the exchanges between blood and the brain occur (Fig. 16a). Along this network, red blood cells that bind oxygen through hemoglobin will meet the increased demand for oxygen of active neurons. Importantly, one specificity of the brain is that capillary endothelial cells are highly selective and semi-permeable, preventing blood from crossing the barrier of the central nervous system. Indeed, this border, known as the Blood-Brain Barrier (BBB), only allows a selection of molecules to cross through specific junctions/transporters (e.g.: O₂, hormones, ions, glucose, amino acids, etc.) (Fig. 16b).

Importantly, the magnetic properties of hemoglobin change between oxygenated and deoxygenated states. The level of blood oxygenation, also called the blood oxygenation level dependent (BOLD) signal, is what fMRI measures to infer neural activity.

1.2 Physics

The MRI scanner is simply a huge cylindrical magnet that emits a powerful electromagnetic field. A typical field strength is 3 Tesla (T), that is equivalent to approximately 50 000 times the strength of the Earth's magnetic field. Some scanners induce stronger fields (currently, an increasing number of human neuroimaging studies work at 7 T, but scanners with stronger fields are beginning to see use), resulting in images with higher spatial resolution and quicker examination time on average.

The electromagnetic field (i.e.: emission of electrons) generated by the scanner affects the spatial arrangement/spin of protons (H⁺) contained within the atomic nuclei of water molecules (H₂O), both within and around the scanner. As H⁺ absorb electromagnetic energy (when exposed to short radio frequency pulses), they tend to align in the direction of the magnetic field. Once they release the energy, they return to the initial state (random orientation/direction). The energy released during relaxation produces a global coherent magnetic signal large enough to be measured by the head coil. Varying the strength of the field and the frequency of radio frequency pulses affect the degree of alignment and the relaxation time, hence resulting in a characteristic signal (Fig. 17). A key property of this magnetic signal is that it varies in strength depending on the material through which it travels. For example, the signal echoing from grey matter, white matter, or cerebral spinal fluid (CSF) have different relaxation times and, depending on the type of MRI sequence used, will show specific contrasts.

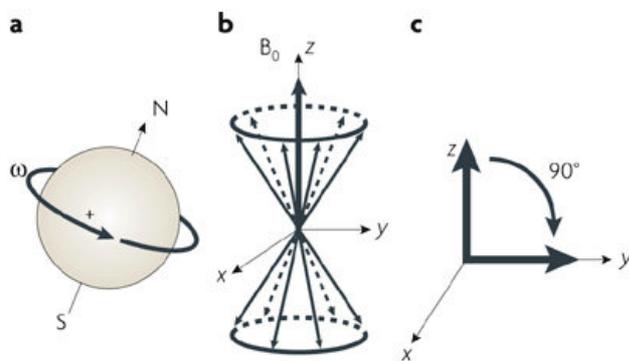


Figure 17: Illustration of the magnetization vector (Brindle, 2008)

Some nuclei have the property of spin, meaning that they possess a magnetic dipole moment (a). When placed in a strong magnetic field (B_0) these dipoles tend to align with the field, although not precisely (b). Application of an oscillating magnetic field that is perpendicular to the main field (B_0) and rotating coherently with the nuclear spin induces transitions between two energy levels and tips the net magnetization vector into the x - y plane, hence resulting in a detectable signal in the receiver coil. The second magnetic field is usually applied in the form of a RF pulse (e.g., 90° , c).

To induce specific image contrasts that provide complementary information (about the brain's anatomy, function, etc.), various MRI sequences are implemented, with highly controlled technical parameters. Two main temporal settings that govern the properties of the resulting images are Echo Time (TE) and Repetition Time (TR), both measured in milliseconds. TE corresponds to the time between the center of the radio frequency pulse to the center of the generated echo signal. TR is the time between two corresponding consecutive points in a series of pulses and echoes.

Structural (or T1-weighted) images

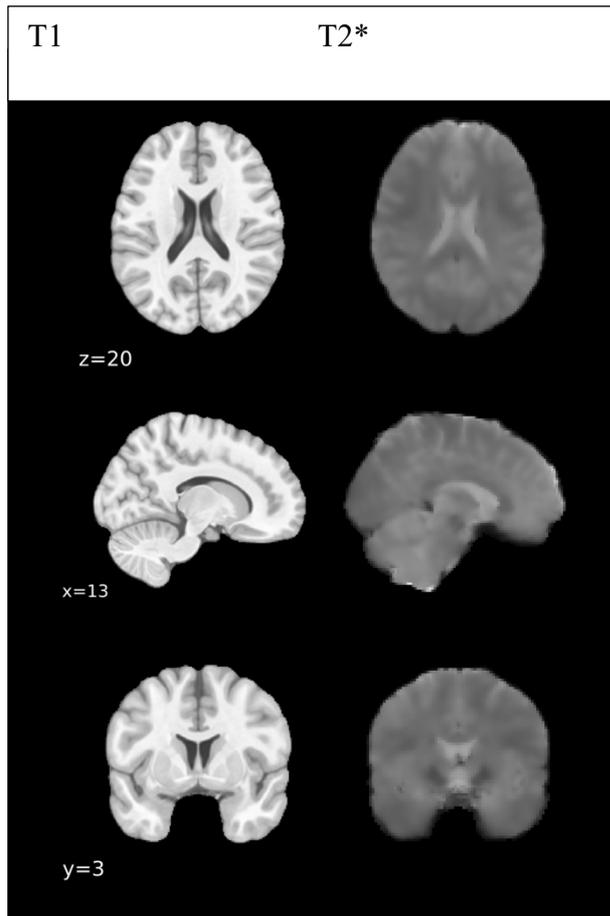


Figure 18: Anatomical (T1) and Functional (T2) images*

Structural MRI sequences (or T1 images in Fig. 18) are characterized by both short TE and short TR. This type of sequence is used to describe the anatomy of the brain, both cortical surface and volume, in terms of shape, size, composition and organization. Notably, the contrast and the difference of relaxation time between white matter and grey matter is determined by fundamental anatomical differences. White matter is mainly composed of axonal projections covered with myelin, a fatty substance that protects and sheathes axons, and very few cell bodies (mainly glia). Therefore, it is less dense to electrons than grey matter, that is heavily filled with neurons and glial cells. For this reason, on a T1w image, white matter appears particularly bright, whereas grey matter is darker, and CSF is not detected.

Functional (or T2-weighted) images

When aiming to investigate the brain's functional activity, another family of MRI sequences is used, referred to as T2-weighted. Traditionally, these sequences show long TR and TE and have the specificity to delimit brain tissue (dark) from CSF (bright). In addition, some T2w sequences ensure a contrast between white (visualized as dark grey) and grey matter (as light grey).

In contrast to T2w sequences that trigger single spin echo signals, an alternative approach consists in inducing a gradient echo train (i.e., series of echoes that are generated in response to each radio frequency pulse), referred to as T2*w (T2* in Fig. 18). This property can be explained by T2*w sequences showing very short relaxation time ($TR_{T2^*w} < TR_{T2w}$). The main advantage of such a sequence is its efficiency, with quantitative data being acquired over a short period (ensuring better temporal resolution, ~ 2 sec). However, this comes at the expense of spatial resolution (increased effect of susceptibility distortions due to field inhomogeneities, mainly around air/tissue interfaces). Nevertheless, such artefacts can be corrected by running a reverse phase-encoded gradient sequence (Fig. 19).

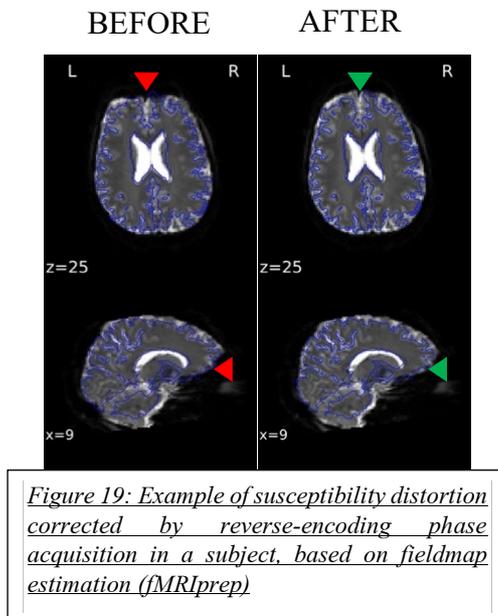


Figure 19: Example of susceptibility distortion corrected by reverse-encoding phase acquisition in a subject, based on fieldmap estimation (fMRIPrep)

Gradient Echo-Echo Planar Imaging

Gradient Echo - Echo Planar Imaging (GE-EPI) is the most popular type of functional MRI sequence because it constitutes one of the fastest imaging techniques (with approximately 1 slice acquired every 100 ms). The signal is acquired almost continuously under the form of a gradient echo train (i.e., a series of echoes generated in response to each radio frequency pulse). This is implemented by rapid alternation of phase and frequency encoding gradients.

MultiBand accelerated EPI pulse sequences

Multiband accelerated EPI refers to ultrafast image acquisition with simultaneous excitation and acquisition of multiple slices per radio frequency pulse (instead of standard single slice stimulation). This method is particularly adapted for event-related protocols that require an optimized temporal resolution (TR~1 sec). However, it requires an "unfolding process" to correct for aliasing artefacts and to increase Signal-to-Noise Ratio (SNR). This allows for a significant reduction of the effective TR by the number of simultaneously stimulated slices (depending on the acceleration factor).

2. Modeling BOLD signal with the General Linear Model

2.1 Introduction to GLM statistics

The first statistical model of importance when dealing with fMRI data is the General Linear Model (GLM), a tool that enables estimating condition-specific contrasts in comparison to other conditions and/or baseline activity, while generalizing across events/trials, functional runs, sessions, observers etc. The main purpose of such a model is to test if a significant variation of BOLD signal can be attributed specifically to a variable/condition of interest (i.e., "Is the difference in BOLD estimates significantly different from zero between conditions A and B?").

The GLM is a linear model and can be explained with reference to linear regression:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \varepsilon$$

Where y corresponds to the measured BOLD signal,

x the predictor variable(s)

$\hat{\beta}_0$ an intercept term,

$\hat{\beta}_1$ estimate(s) of the slope or explanatory weight,

ε the residual error* (i.e., remaining unexplained variance, presumed to follow a Gaussian distribution with zero mean and variance of σ^2)

*Typically, it is important that quantitative fMRI data is acquired repeatedly for each condition of interest and baseline, so that the residual error for the mean is normally distributed, following the Central Limit Theorem.

The GLM is used to model many repeated measures in time (time series). Therefore, the response variable y as well as the predictor x and the error ε are expressed as matrix equations as follows (1), usually rearranged into a matrix multiplication (2):

$$(1) \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \end{bmatrix} .$$

$$(2) \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ 1 & x_5 \\ 1 & x_6 \\ 1 & x_7 \\ 1 & x_8 \\ 1 & x_9 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \end{bmatrix} .$$

In addition, the GLM design matrix will generally include more than one explanatory variable. Therefore, the more regressors added (experimental or control conditions, button press, confounding factors such as head motion etc.), the more columns in the design matrix and the more complex the model.

In design matrix (3), the first column of 1s represent the constant or baseline activity, whereas the three following correspond to different experimental conditions that can be compared via GLM contrast estimates. The corresponding β_i coefficients estimate the weight of each explanatory variable x_i in describing the measured y_i signal.

$$(3) \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & x_1 \\ 1 & 1 & 0 & 0 & x_2 \\ 1 & 1 & 0 & 0 & x_3 \\ 1 & 0 & 1 & 0 & x_4 \\ 1 & 0 & 1 & 0 & x_5 \\ 1 & 0 & 1 & 0 & x_6 \\ 1 & 0 & 0 & 1 & x_7 \\ 1 & 0 & 0 & 1 & x_8 \\ 1 & 0 & 0 & 1 & x_9 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \end{bmatrix} .$$

For further information, Poline and Brett (2012) published a very clear and concise review on GLM methods and applications, especially for fMRI.

2.2 Multilevel GLM analyses

Subject vs Group-level statistics

In the previous section, a very schematic description of the GLM was provided. Nonetheless, statistically comparing GLM contrast estimates across multiple functional runs, sessions, or between individuals requires the implementation of several levels of analysis. Initially, one design matrix per run is generated, then a second-level design matrix at the session/individual level, and finally statistical analyses are pursued up to the group-level if a common standard template is used (e.g.: MNI152NLin2009cAsym).

ROI-based analyses

Defining regions of interest (ROI) that are hypothesized to be specifically involved in the investigated function is a way to minimize the issue of multiple comparisons (i.e., testing independently a large number of GLM contrasts at the voxel-level, which increases the Type 1 error rate). However, certain rules must be followed in order to ensure that the localized ROI is not biased in relation to the contrast of interest (e.g., arising from non-independence of measurements, see Kriegeskorte et al, 2009, 2010). For example, if a functional “localizer” sequence is used for the voxel selection of the ROI, then, the contrast of interest for the main experiment should not overlap with the localizer sequence. Moreover, ideally, each ROI analysis should be treated as an independent test.

Regressing out confounding factors

BOLD signals result from both neural and non-neural variables. As previously described, neural activity is inferred from the local changes in the BOLD signal. Non-neural fluctuations, however, arise from additional confounding sources: head motion, physiological fluctuations (due to cardiac or respiratory effects), etc. Greve et al., 2013 reviewed many of the potential sources of noise occurring during fMRI data acquisition (Fig. 20). If these potentially confounding signals are measured, they can be included as covariates in the GLM design matrix, to estimate their influences on the signal and thereby reduce their contribution as a source of uncontrolled variation or noise in interpreting the relevant contrasts.

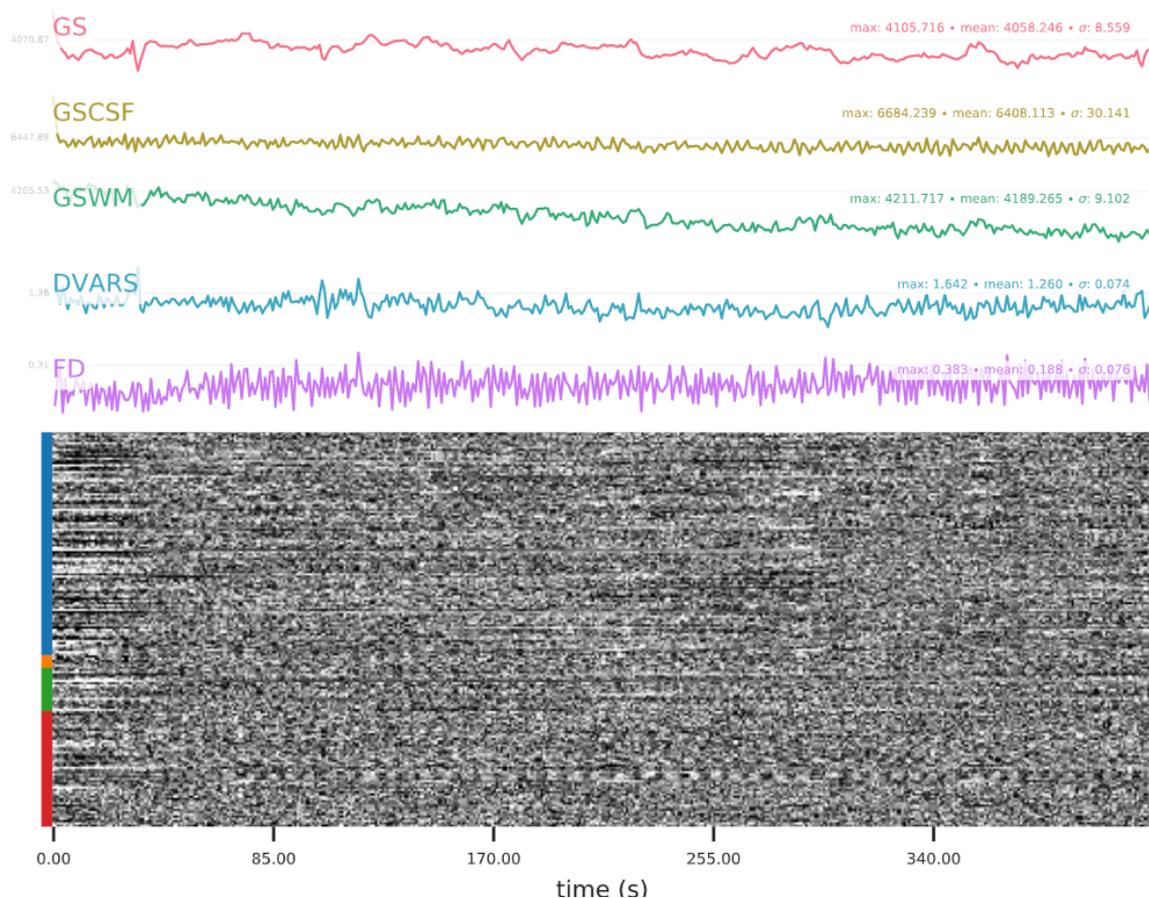


Figure 20: Example of confounds that can be derived from global signal (with fMRIPrep)
Anatomical: GS: global signal, -CSF: Cerebro Spinal Fluid, -WM: White Matter
Head motion: DVARS is the derivative of the estimated relative frame-to-frame bulk-head motion over voxels (Jenkinson et al., 2002).
FD Framewise Displacement estimates bulk-head motion calculated using formula proposed by Power et al., 2012ab)

3. Material & Method

Participants

Thirty-four healthy participants (24 females, mean age 25.5, s.e.m. 0.61) with normal or corrected-to-normal vision were included in the study. They each performed both a localizer task and a bistable motion experiment. Three participants were excluded because of poor performance on the fixation task and/or in reporting their perceptual state over time. All participants gave informed consent and were paid 20 euros for their participation.

The experimental procedure was covered by the CMO regio A-N 2014/288 blanket approval called “Imaging Human Cognition” between the Donders Center for Cognitive Neuroimaging (DCCN) and the Medical Research Ethics Committee.

Experimental design and apparatus

Before entering the scanner, each participant was allowed to practice and become familiar with the task during a short behavioral training period. The experimental session included simultaneous fMRI, behavioral and eye movement recordings. The scanning session included the acquisition of a standard structural MRI sequence (T1 MPRAGE 1mm ISO), together with a head scout and localizer (32-channel head coil). Functional volumes were Multi-Band accelerated Echo Planar Imaging (EPI) Pulse sequences (MB6 2mm ISO, TR1000). This type of accelerated sequence has the advantage of using a short repetition time (1000 ms) and a 6-fold acceleration factor (number of slices simultaneously excited) that results in the acquisition of a multiband signal in a single EPI echo train. To correct for EPI image distortion, we also performed a field map and an inverted phase encoding direction EPI sequence.

During the bistable motion task, and while fixating a central cross, observers were presented with a bistable moving plaid. Their task was to report dynamically each perceptual switch of motion percept they experienced. A three-alternative choice was made possible by means of a button box, and observers indicated their perceptual state at a specific time: component motions (“Component”), plaid (“Pattern”), or, if they were uncertain, an ambiguous percept (excluded from the analyses, see Analysis pipeline – behavioral data). In total, they performed 5 runs, with six one-minute trials each, of this task interrupted by 10 second intervals of intertrial break. The Component/Pattern report of perception served as a subjective measure reflecting perceptual decisions. The transition rate between the two perceptual states, i.e., how frequently subjects’ perception alternated, was used as a measure of motion bistability.

In the same session as the main bistable perception task, 3 runs of functional localization were achieved. This localizer had two main purposes: first, defining the bilateral motion-selective complex hMT+ ; secondly, selecting within each visual area direction-selective subdomains.

Main stimulus: bistable plaid

The bistable plaid stimulus was tested and described in detail in the Psychophysics experiment (see Chapter 2, Fig.10 reported below).

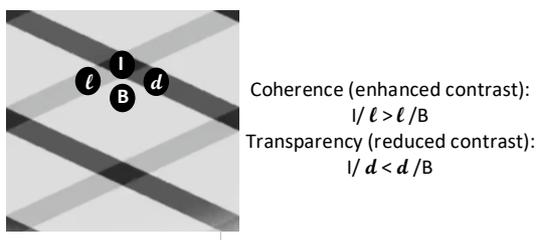


Figure 10: Transparency versus Coherence, a matter of depth-order
Component motion perception is facilitated by transparency between gratings. A way to implement this is to vary the contrast of intersection region (I) relative to the rest of the plaid (l, d, B). Reducing this contrast will more probably elicit the component percept, whereas enhancing it has the opposite effect (pattern). For this reason, we created a hybrid plaid with one grating in reduced, the other in enhanced contrast regime (Stoner & Albright, 1990; 1992; 1996; 1998).

A critical experimental design feature necessary for inducing and measuring bistable motion perception is long-term viewing of the ambiguous stimulus (for review, see Hupé & Rubin, 2003). For this reason, continuous one-minute trials were performed in this experiment.

Long term presentation of the plaid stimulus resulted in an ambiguous motion perception, alternating between two percepts that remained stable for several seconds at a time. We will refer to these as Component (C) and Pattern (P) perceptual states. The stimulus input being

unchanged, we considered that each transitory button press corresponded to a perceptual switch towards a new state as reported by the observer.

Localizer stimulus

The functional localizer consisted of short trials (10 seconds) of non-ambiguous moving gratings randomly interleaved with static ones (each type repeated 13 times). The moving gratings were presented specifically in the directions matching either the perceived component (upward and downward directions) or pattern (rightward direction). Disambiguating the movement of lines seen through a small aperture design while preserving the orientation/direction combinations of interest, required modification of the original stimulus. In this context, we chose to use single gratings made of dashed lines, allowing the extraction of information from the local edges of the dashes and thereby restoring a two-dimensional motion signal. We also enhanced the contrast of the stimulus (white lines moving on a black background). This resulted in a strong and clear perception of motion direction (Fig. 21).

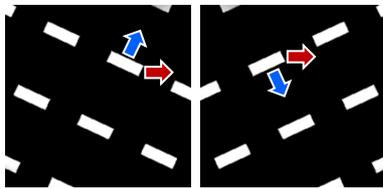


Figure 21: Direction-selective localizer stimulus with disambiguated moving gratings

Dashed-line gratings moving either obliquely (blue arrows) or horizontally (red arrows) were shown to participants during 10-second trials. Gratings orientation and direction matched those perceived in the bistable plaid stimulus.

Based on this independent localizer, direction-selective GLM contrast estimates enabled to define direction-selective units within visual cortical areas.

Imaging data acquisition

All MRI data were acquired on the 3T MAGNETOM Skyra MR scanner (Siemens AG, Healthcare Sector, Erlangen, Germany) at the Donders Center for Cognitive Neuroimaging (Nijmegen, Netherlands), using a product 32-channel head coil. All participants performed a unique fMRI scanning session of approximately 1 hour. This session was composed of a standard MPAGE T1w anatomical scan (315 s, resolution 1mm³, during which a training was performed), 5 runs of bistable motion (each lasting 430 s) and 3 runs of motion/direction localizer (400 s each).

The anatomical scan consisted of a T1-weighted Magnetization Prepared Rapid Acquisition Gradient Echo (3D-MPAGE, 256 sagittal slices, TR = 2300 ms, TE = 3 ms, flip angle = 8°). Furthermore, an initial slice positioning fieldmap was collected to ensure the functional coverage of V1, V2 and hMT+ (66 coronal slices, isotropic resolution of 2 mm).

During the experimental runs, the BOLD signal was recorded using MultiBand (MB) accelerated Echo Planar Imaging (EPI) sequences with an acceleration factor of 6 (66 coronal slices, isotropic resolution of 2 mm, TR = 1000 ms, TE = 34 ms, Anterior-Posterior phase encoding). Finally, due to the fact that MB accelerated sequences provide a better temporal resolution at the cost of higher susceptibility-related artifacts, EPI data with the opposite phase gradient (Posterior-Anterior) were collected to efficiently correct for gradient field non-linearity distortions.

4. Analysis pipeline

4.1 Behavioral data

Participants indicated each perceptual switch by using a bimanual 8-button box response pad (HHSC-2x4-C). Three alternative choices were proposed, either reporting the Components, the Pattern, or reporting a mixed transitory percept. In order to avoid confusion and errors in the button presses, and to optimize response time, we assigned Component/Pattern responses to different hands (L/R). The subjective reports were used to define experimental conditions and study temporal dynamics of bistable perception, and later to realign and analyze the fMRI and eye tracker recordings. Any percepts shorter than or equal to 2 seconds or longer than 30 seconds, and any report of mixed percept were considered as outliers and excluded from the analyses (Zhang et al, 2017). Excluded outliers represented on average 20% of the total dataset.

4.2 fMRI Preprocessing

Preprocessing steps were applied following fMRIPrep analysis pipeline. First, T1w images were corrected for intensity non-uniformity (Tustison et al. 2010; Avants et al., 2008) and skull-stripped (Nipype implementation of the ANTs atlas-based brain extraction workflow). Brain tissue segmentation of CerebroSpinal Fluid (CSF), White-/ Grey-Matter (WM/GM) was defined (based on FSL Fast toolkits). Volume-based spatial normalization of the T1w reference was performed to the MNI152NLin2009cAsym standard space through non-linear registration (ANTs registration tool; Fonov et al, 2009).

Regarding BOLD series, a reference volume was computed out of the aligned single-band reference images (SBRefs). A field map estimating susceptibility distortions was estimated based on the reversed phase encoding EPI sequence (AFNI; Cox, 1996) and enabled unwarping the data, providing a corrected EPI reference. This latter was used to co-register functional data to the T1w reference (using `bbregister` in Freesurfer) by implementing a boundary-based registration (Greve & Fischl, 2009) with 6 degrees of freedom. Head-motion parameters were estimated with respect to the functional reference (transformation matrix including 6 rotation and translation parameters, using `mcfliirt` in FSL; Jenkinson et al, 2002) before any spatio-temporal filtering was applied. The BOLD signal was slice-time corrected (using `3dT-shift` in AFNI; Cox, 1996). The resulting BOLD time-series were then resampled to both native and standard space by applying a single composite transform to correct for head-motion and susceptibility distortions.

fMRIPrep calculates several confounding time-series based on the preprocessed BOLD signal. Framewise displacement (FD) and DVARS were calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al., 2014). Three global signals were extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction using the CompCor method (Behzadi et al., 2007). Principal components were estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128 s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are calculated from the top 2% variable voxels within a mask covering the subcortical regions. aCompCor components are calculated within the intersection of the

aforementioned mask and the union of CSF and WM masks calculated in T1w space, after projection to the native space of each functional run.

After preprocessing with fMRIPrep, the confounds were inspected to determine if data met the criteria for inclusion. Subjects were excluded if more than 30% of time points exceeded a FD of 0.5 mm. Another reason for excluding subjects was if they did not experience a bistable perception or did not perform well at fixating the target cross when viewing the ambiguous plaid stimulus. For each subject, confound time-series included in the model were the six head-motion parameters, the first six aCompCor components from WM and CSF mask, and FD.

BOLD data were smoothed using SUSAN noise reduction nonlinear filtering that reduces noise whilst preserving the underlying structure. This is accomplished by averaging only neighboring voxels showing similar intensity, i.e., that form part of the “same region” as the central voxel. A 6 mm FWHM gaussian kernel was applied and the threshold for brightness intensity set to 75% of the median value for every run.

Afterwards, each volume of every run was scaled so that the median value of a specific run was set to 10000, and a scaling factor was estimated for intensity normalization. Finally, BOLD data were band-pass filtered using high-pass (0.008 Hz) and low-pass filtering (3rd-order polynomial function using the method of linear least squares) (see Savitzky & Golay, 1964; Press et al., 2007).

4.3 fMRI analysis: GLM

BOLD responses recorded during both the bistable motion and the functional localizer were assessed by applying an ROI-based fixed-effect general linear model (GLM) approach using the FSL implementation in Nipype. All regressors representing experimental conditions were convolved with a Double-Gamma hemodynamic response function (dgamma-HRF). For the localizer data (3 runs), four predictors (oblique (2) and rightward (1) motions, static stimulus) were represented with boxcar functions. For the bistable motion sequence (5 runs), we modeled the perceptual report (Component, Pattern and or Mixed motion) with Single Impulse (or stick) functions. The first-level (run-level) design matrix included predictors (experimental condition, button press) and motion correction parameters as confound predictors. Voxel timeseries were z-transformed and corrections for serial correlation were applied (using a first-order autoregressive estimator). At this step, any run that did not fulfill the inclusion criteria either because of excessive head motion, poor fixation stability or unclear bistable perception was excluded from the analyses. In total, 3 participants were fully excluded from the study, two of them did not properly perform the task, and one excessively moved his head. Finally, for one participant, we decided to discard two bistable motion runs (out of five), because of excessive head motion (average FD>0.5 mm).

The second-level (or subject-level) GLM design matrix was then created for each of the remaining 34 subjects by merging the run-based contrast parameters. The average BOLD response (z-score) was then extracted for each region of interest (ROI: direction-selective subdomains) within visual areas (V1, V2, and hMT+).

Finally, group-level statistics were calculated by implementing a Linear Mixed-Effects model (Bates et al, 2015) to account for between-subject variability, while testing the effect of variables (or interaction of variables) that best fitted the measured response variable.

Brief introduction to Linear Mixed-Effect Models

Once again, let's start with a formal linear regression model that would estimate the BOLD signal as a function of explanatory variables in a matrix X:

$$Y \sim \beta_0 + X\beta_1 + \varepsilon$$

Where Y is the response variable,

β_0 represents the intercept (estimate of response variable Y when X = 0),

X, the design matrix of predictor variables,

β_1 , vector of regression coefficients (fixed effect),

ε , error accounting for the residual variance (following normal distribution).

This type of linear regression is ideal for homogeneous datasets. However, when there is significant variability within the observed sample due to multiple sources of variance (i.e., a structure including multiple randomly chosen levels, groups or individuals), then Linear Mixed Effect (LME) models will be the appropriate approach to describe the data. Here, the participants were considered to be randomly chosen so that the variance due to the choice of the participants sampled from the population is incorporated and estimated as an additional source of variance in the model.

In the current situation, LME models consist in regression analyses that measure a dependent variable (the response variable) as a function of explanatory variables (or "effects") measured across many repetitions on different participants. There are two main types of effects: fixed and random. A fixed effect will be used to test the effect of an explanatory variable of interest X (or interaction of variables) on response variable Y (with the variable levels of interest being set by the experimenter). On the other hand, a random effect applies to any variable whose level is considered a random draw from the population of potential levels that could have been chosen for the experiment. In a typical case, the random choice of subjects who participated in the experiment fits this definition. Only the variance of the random effects is estimated directly in fitting the model. The LME approach is widely used in experimental designs involving multiple levels of analysis (e.g., within-subject/group-level analyses). In fMRI, a very common application is for considering the effect of parameter X on the BOLD signal both at subject- and group-levels. In that case, several sources of residual variance in the data can be considered as potential random effects: subjects, sessions (if there were multiple scanning per participant), functional runs, etc. In addition to fitting the data better, this has the advantage of enabling more accurate predictions when aiming to generalize conclusions beyond the scope of the dataset.

An example of a simple LME model that includes a random intercept term is the following:

$$(Y|sub) \sim \beta_0 + \beta_1 X + \gamma_{0,sub} + \varepsilon$$

Where Y is expressed conditionally on random effect sub,

β_0 represents the fixed or population intercept (estimate of Y when all independent variables are null),

X, the design matrix of predictor variables,

β_1 , vector of fixed effect coefficients,

$\gamma_{0,sub}$, a subject specific random intercept, the product of the design matrix of random intercepts and a vector of random-effect coefficients that is assumed to be normally distributed and whose variance rather than the individual coefficients is estimated in the model,

ε , error accounting for the residual variance (normally distributed)

The LME model can be made more complex as more terms are added, effects of interaction, or random slope coefficients (i.e., when the random effect includes variance terms for different intercepts and slopes per subject, for example). However, in order to avoid overfitting the data, the trade-off between goodness of fit and model complexity is taken into account.

Regions of interest (1): visual areas

The Freesurfer reconstruction tool enabled delineating anatomical regions of interest bilaterally, namely retinotopic visual areas V1 and V2, as well as hMT+ complex, the contiguous region of gray matter in the posterior middle temporal region responding more strongly to moving gratings than to stationary ones (selection of the 100 most specific voxels from the [Motion – Static] GLM contrast, $p < 0.01$). For each participant, we first extracted bilateral V1, V2 and hMT+ masks (average size \pm sd, isotropic resolution 1mm: 9946.6 ± 753.1 voxels ; 17037.0 ± 953.9 voxels ; 4060.5 ± 388.2 voxels, respectively).

Within these anatomical masks, we used a functional localizer to reveal direction-selective populations of voxels.

Decoding perceptual state

In a first attempt to identify the neural codes of Component and Pattern perceptual states, we performed subject-level classification analyses. Given the rapid event-related design paradigm, we chose to use single-event beta estimates with the Least Squares Single (LSS) method (described in detail in Mumford et al., 2012; 2014). This consists of modeling a betaseries for each trial or event (a perceptual state for instance) independently in each voxel, instead of defining one average regressor representing a trial type as is done in more standard approaches. The advantage of such method is that it enables deconvolving the BOLD activity, particularly in the case of adjacent events that we want to distinguish. We implemented this using the NiBetaSeries tool (<https://nibetaseries.readthedocs.io/en/stable/usage.html>), with the LSS estimator and Glover's Double-Gamma hemodynamic response function. Similar to the preprocessing steps described in a previous section of the current chapter (4.2 fMRI Preprocessing), a 0.008 Hz high-pass filtering was applied and the same motion regressors were added to the model. However, specifically for this analysis we did not perform any spatial smoothing that could interfere with the multivoxel pattern analysis (Misaki et al., 2013). Furthermore, to prevent potential confounding factors (e.g., behavior, oculomotor activity, etc.)

biasing the classification accuracy, we used independent datasets to train and test logistic regression classifiers to discriminate motion direction for each subject. After being normalized at the run-level, the betaseries were extracted and labeled according to their trial type (i.e., Rightward or Oblique for the localizer; Pattern or Component for the bistable motion task). The training step was carried out entirely within the localizer dataset, classifying multivoxel patterns of activity between Rightward and Oblique conditions. Subsequently, [Rightward vs Oblique] trained classifiers were tested at decoding the matching perceived directions in the bistable motion experiment, thereby aiming to discriminate Pattern from Component perceptual states. The goal of this analysis was to evaluate how similar neural representations of motion direction were between the stimulus-driven responses and the bistable perception.

In order to assess the prediction accuracy of the model, we used the Receiver Operating Characteristic (ROC) curve, a metric that describes the trade-off between the true positive rate (TPR) and false positive rate (FPR) of a prediction. The prediction accuracy of the model can be assessed by the average area under the curve (AUC). The higher the AUC score, the more accurate the classification. According to David W. Hosmer and Stanley Lemeshow (2004), AUC scoring can be evaluated as follows:

- 50 % = No discrimination (random)
- 50-70 % = Poor discrimination
- 70-80 % = Acceptable discrimination
- 80-90 % = Excellent discrimination

At the group-level, average classification accuracies are represented in figure 22 for each functionally defined motion-selective region in V1, V2 and hMT+. For each region, classification accuracy was compared to chance-level based on $n=1000$ permutation tests (i.e., after randomizing trial type labels of events). At the group-level, in V1 and V2, the from-localizer-to-bistable classification analysis failed in generalizing decoding to perceptual states (unilateral paired t-test; V1: $t=-1.54$, $df=33$, $p=0.85$; V2: $t=-2.82$, $df=33$, $p=0.99$). Although classification was significantly more accurate in hMT+ ($t=2.95$, $df=33$, $p<0.01$), the classification accuracy remained within the range of poor discrimination (54.1%) (Fig. 22). This suggests that neural representations of internally induced perceptual states are dissimilar to those reflecting physical variations of motion signals. We hypothesized that such observations may indicate that neural units triggering perceptual switching are specific to top-down induced percepts.

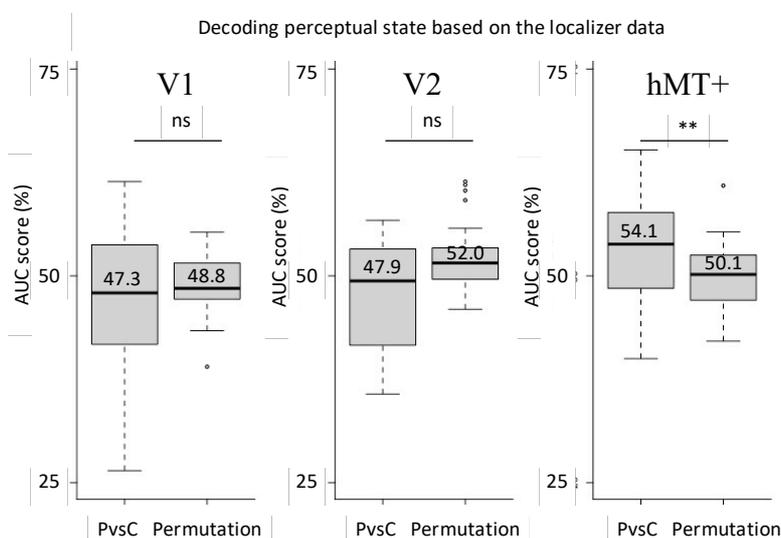


Figure 22: Generalized localizer-to-bistable classification in motion-selective visual regions

Average z-scored single-event estimates were extracted. Logistic regression classifiers were trained on the localizer dataset to identify multivoxel patterns of activity representing motion directions [Rightward vs Oblique]. These classifiers were tested on the bistable motion dataset, decoding [Pattern vs Component] (PvsC) perceptual states. Group average AUC scores of classification were compared to chance level ($n=1000$ permutation tests).

Classification was significantly above chance level only in hMT+ (**, $p<0.01$), and not significant in V1 and V2 (ns). $N=34$ observers

These results motivated the need to examine the activity of regions of interest which are more selective to the feature of interest, that is perceived motion direction, and to compare locally BOLD signal variations relative to perceptual states. Therefore, we decided to investigate bistable motion perception from the perspective of direction-selective units.

Region of interest (2): Localization of direction-selective subdomains

Each area's average response to the unambiguous localizer sequence was used to map direction-selective subdomains independently of the bistable perception experiment. Prior to determining the optimal subdomain size, we considered how generalizable direction-tuning was across localizer runs. To evaluate this, we used a 3-fold cross-validation approach where each iteration consisted of two steps: 2 runs were picked to select the N most direction-selective voxels, and the remaining run was used to test the specificity of the selection in the estimated GLM contrast [Rightward vs Oblique]. This leave-one-run-out approach was repeated for different subdomain sizes (N = [25, 50, 75, 100] voxels). This range of sizes was determined based on the smallest functionally defined region hMT+ (i.e., after contrast estimate [Motion > Static]) that is significantly smaller than V1 and V2. Quantitative analyses of optimization of unilateral hMT+ functional localization reported an average size that is equivalent to 100 to 150 voxels (for a voxel size of 2mm³) (Huang et al, 2019). Therefore, the maximal subdomain size for both Rightward and Oblique conditions was set to be smaller or equal to half of hMT+ minimal surface (i.e., 50 voxels per hemisphere, or 100 in total), and superior to ¼ of the total surface to ensure sufficient signal-to-noise ratio (>25 voxels). Statistical significance across sizes was evaluated using a LME model (with observer treated as a random effect (Bates et al, 2015)).

$$(Y|sub) \sim \beta_0 + \beta_1 \text{Area/Size/Subdom} + \gamma_{0,sub} + \varepsilon$$

Where Y represents the average z-score of the Rightward vs Oblique GLM contrast estimate conditional on the subject random effect,

Area/Size/Subdom is the nested fixed effect of Subdomain (Subdom: Rightward- or Oblique-selective) within size parameter (Size: 25,50,75,100) within each visual area (Area: V1, V2, hMT+).

The diagnostic plots in Fig. 23 for the model fits indicate that the model assumptions are reasonably met. The figure on the left indicates that there is no evidence of inhomogeneity in the distribution of the residuals nor of systematic bias as a function of the fitted. The normal quantile-quantile plot on the right displays the sorted residuals as a function of the theoretical quantiles of a normal distribution. The alignment of the points with the thin diagonal line indicates that ordered residuals do not deviate systematically from their expected values under the assumption of Gaussian error.

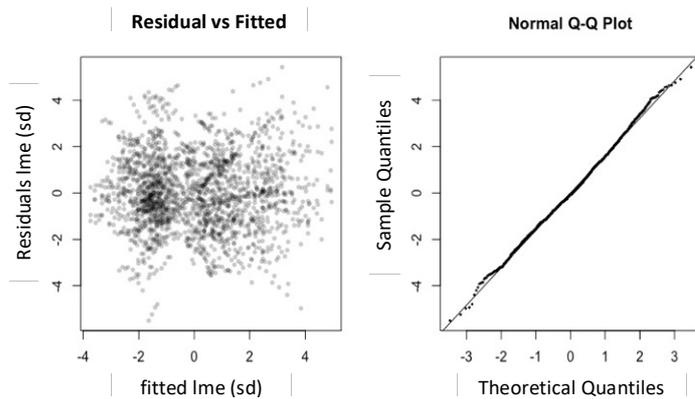


Figure 23: Diagnostic plots of LME model for direction-selectivity (cross-validation)

The Residual vs Fitted plot shows a random distribution of residual variance around zero.

The Q-Q probability plot shows that data follows Normal distribution.

The fixed-effect estimates from the model fit are plotted in Fig. 24 as a function of subdomain size. The z-score decreases approximately linearly with number of subdomain voxels. Thus, direction-selectivity as indicated by z-score was maximized for $N = 25$ voxels in V1 and V2 (the null hypothesis of no difference in selectivity explained by subdomain size was rejected at $p < 0.001$) (table 2). In hMT+, although the differences between voxel size did not reach significance, we observed the same tendency, with $N=25$ voxels ensuring the strongest direction-selectivity.

Therefore, within each area, we selected the top 25 rightward- and oblique-tuned voxels based on the average direction-selective contrast estimate (subject-level GLM, all localizer runs). Overall, these subdomains showed high direction-selectivity with z-scores ranging from about 3.5 to 7.2 as presented in table 2. These subdomains, then, were used as the regions of interest for further group-level analyses.

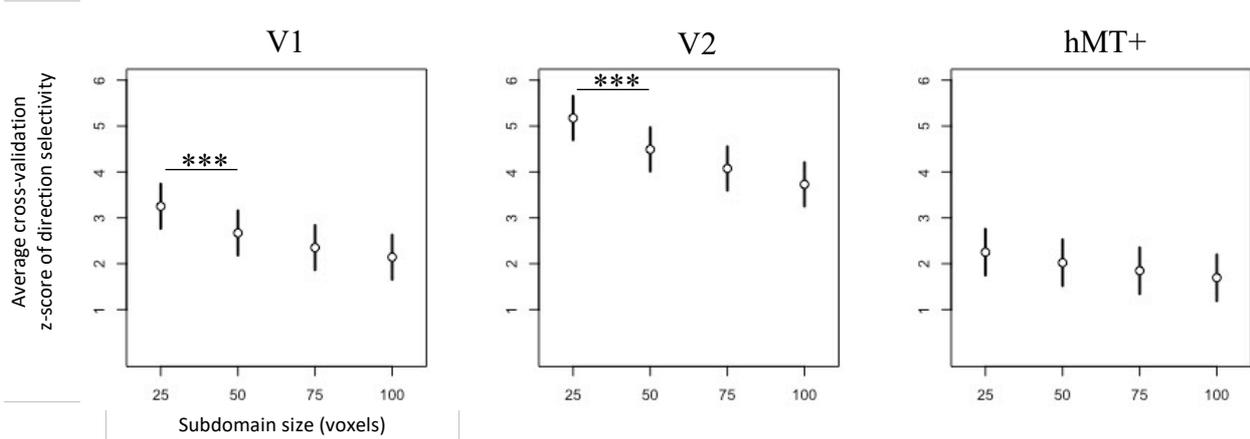


Figure 23: Direction-selectivity of subdomain in visual cortex as a function of subdomain size. From left to right, the coefficients estimated after the cross-validation, investigating the effect of subdomain size on the contrast of direction selectivity [Right vs Oblique] for size = 25, 50, 75 and 100 voxels.

For V1 and V2, the smallest size (25 voxels) maximized direction selectivity within this cross-validation test, showing high specificity (V1: $Z = 3.25$; V2: $Z = 5.17$, $p < .01$)
 In hMT+, although the effect of subdomain generalizing across runs was less strong, and independent of the size, it reached significance (MT: $Z = 2.25$, $p < .01$)
 N= 34 observers (minus 5 excluded due to weak subdomain's localization)
 Fixed effect of subdomain size: ** $p < .01$

Rightward Oblique

	V1	V2	hMT+
Rightward	$Z = 4.80 \pm 1.94$	$Z = 7.25 \pm 2.74$	$Z = 3.54 \pm 1.73$
Oblique	$Z = -5.21 \pm 1.61$	$Z = -5.84 \pm 1.81$	$Z = -4.45 \pm 2.09$

*Table 2: Table of average direction selectivity z-score
 Rightward > Oblique contrast estimate shows high specificity regarding Rightward (z-score > 0, $p < .001$) and Oblique (z-score < 0, $p < .001$) subdomains across visual areas at the group-level
 N=34 observers (minus 5 participants excluded due to weak subdomain localization)*

Investigating perceptual decisions within direction-selective subdomains

For each observer, we estimated the [P > C] contrast (perceptual state) within Rightward and Oblique subdomains. Our main hypothesis was that the P state would preferentially elicit response in Rightward subdomains, whereas the C state would more greatly activate Oblique subdomains. In order to test this hypothesis (against the null hypothesis of no preferential activation of subdomains with percept), we mapped [P > C] contrasts onto V1, V2 and hMT+ and extracted contrast estimates within direction-selective subdomains.

Direction-selective and Perceptual decision maps are presented for five subjects as an example (Fig. 25, see others in Supplementary Fig. S1) associated with a table of z-scores (session-level) per subdomain per area and a caption reporting the main results. The first table reports the average z-values measured in V1/V2 for the contrast estimates of [Rightward > Oblique] (column 1) and [Pattern > Component] (column 2), in both Rightward (1st row) and Oblique (2nd row) subdomains. A similar table is reported below for hMT+. Globally, the first column shows the degree of direction selectivity of the two subdomains, therefore we expect to find $Z > 0$ for Rightward subdomain and $Z < 0$ for Oblique subdomain. If direction-selective subdomains reflect perceived (illusory) motion, we expect the same trend in the second column representing Pattern/Component selectivity ($Z > 0$ in Rightward subdomain and $Z < 0$ in Oblique domain). For 5 subjects (out of 34), we were not able to identify Rightward and Oblique subdomains, therefore we excluded these subjects from group-level analyses and, thus, no z-score tables are included for those participants in Fig. S1.

Imaging data analyses were achieved via Nipype v1.8.2 interface (combining SPM, FSL and Freesurfer modules), and all statistical analyses were performed using the lme4 package (Bates et al., 2015) in the R statistical environment in RStudio (v2022.07.1+554).

S1

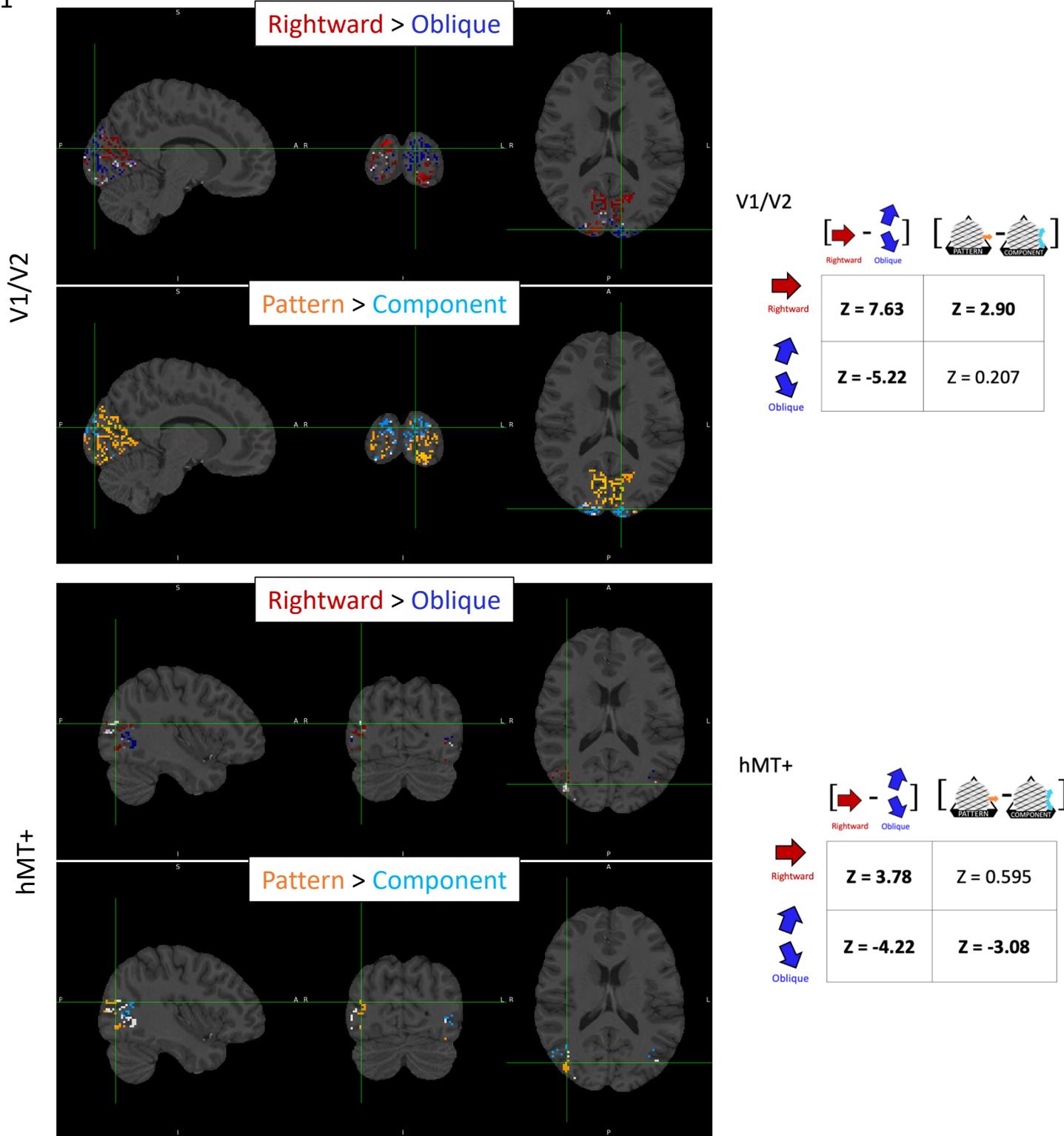


Figure 25.1: Direction-selectivity and Perceptual decision maps in subject 1
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (in blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (in cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

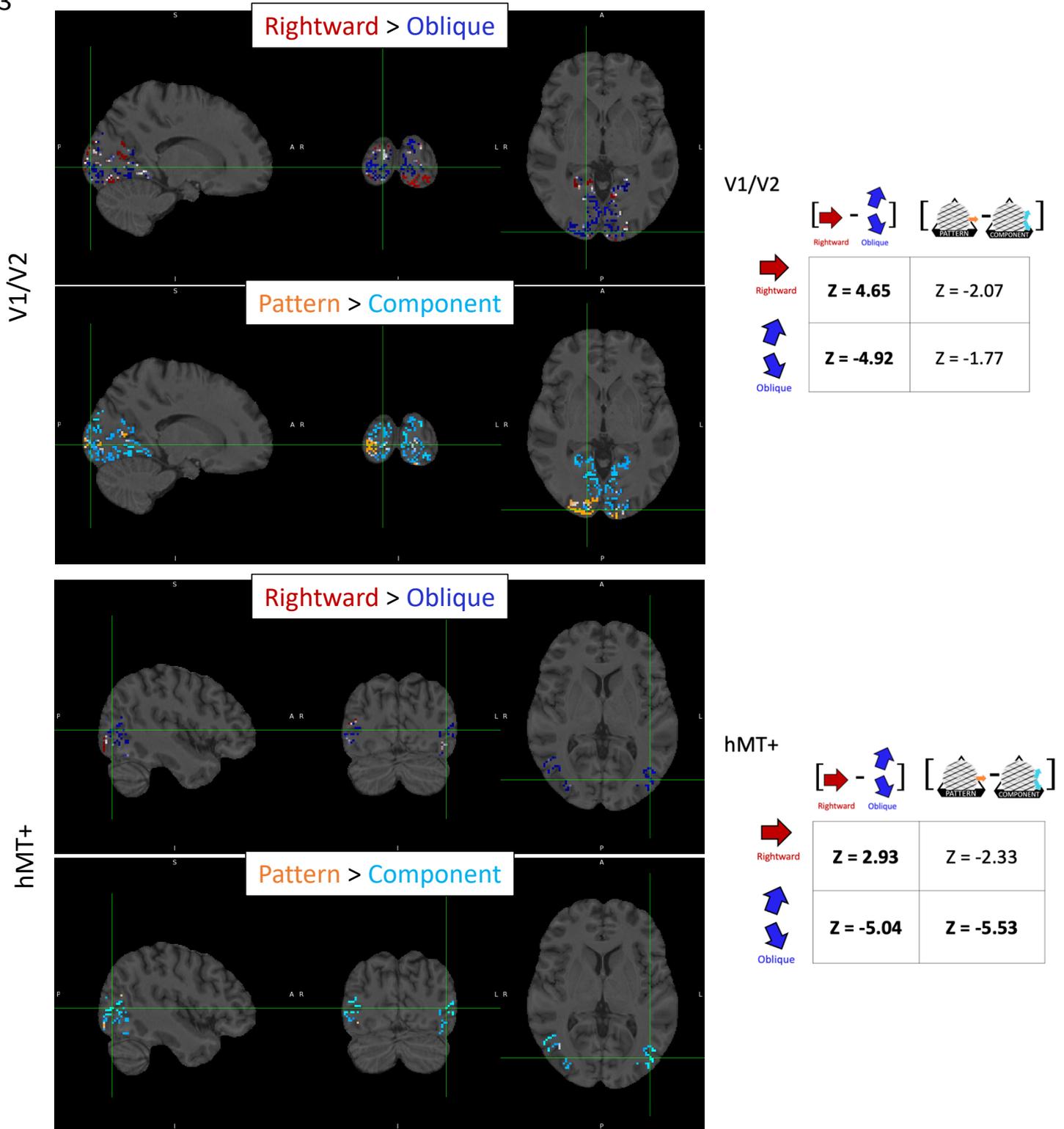


Figure S25.3 : Direction-selectivity and Perceptual decision maps in subject 3
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S4

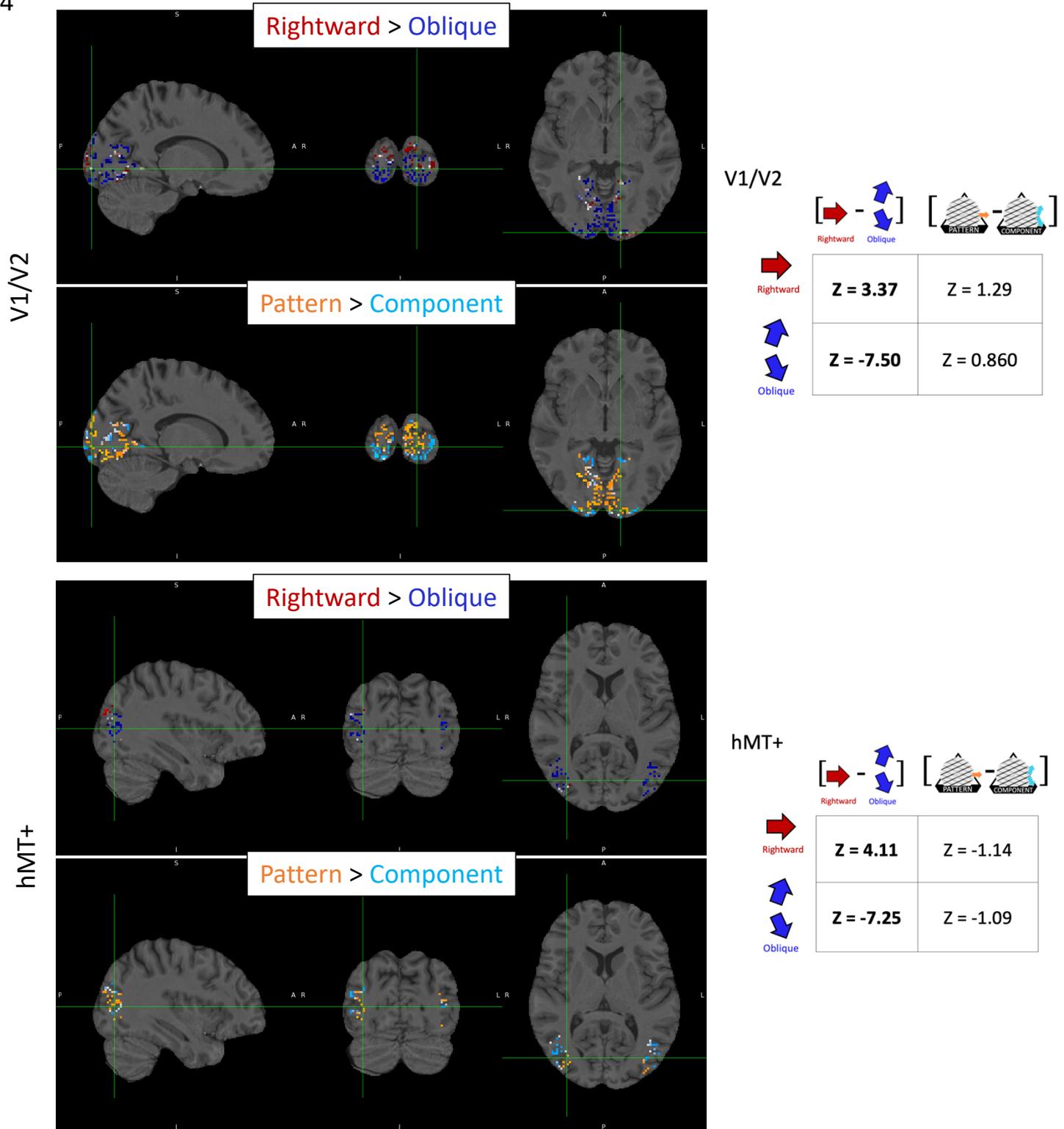


Figure 25.4 : Direction-selectivity and Perceptual decision maps in subject 4
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S5

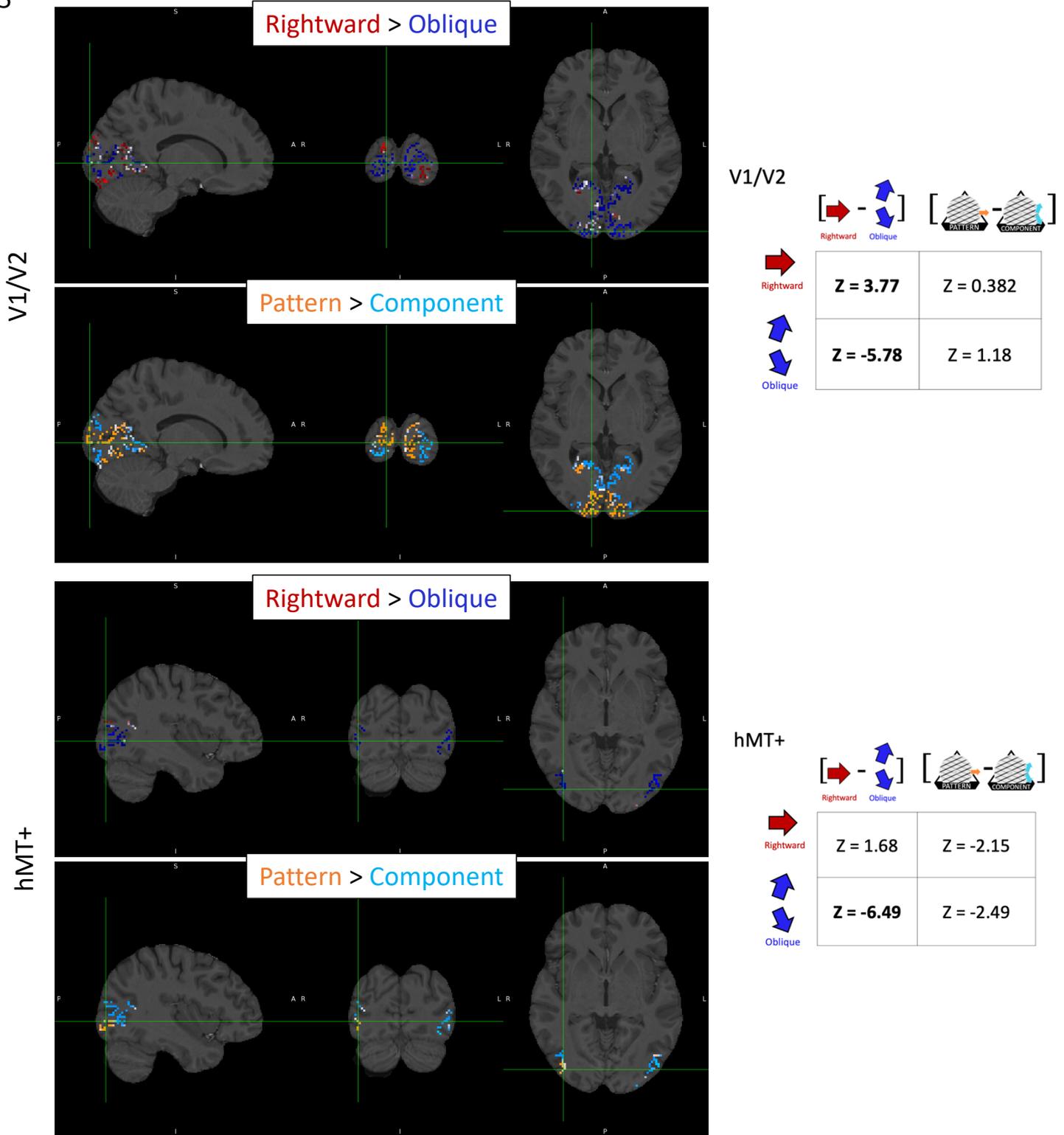


Figure 25.5 : Direction-selectivity and Perceptual decision maps in subject 5
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

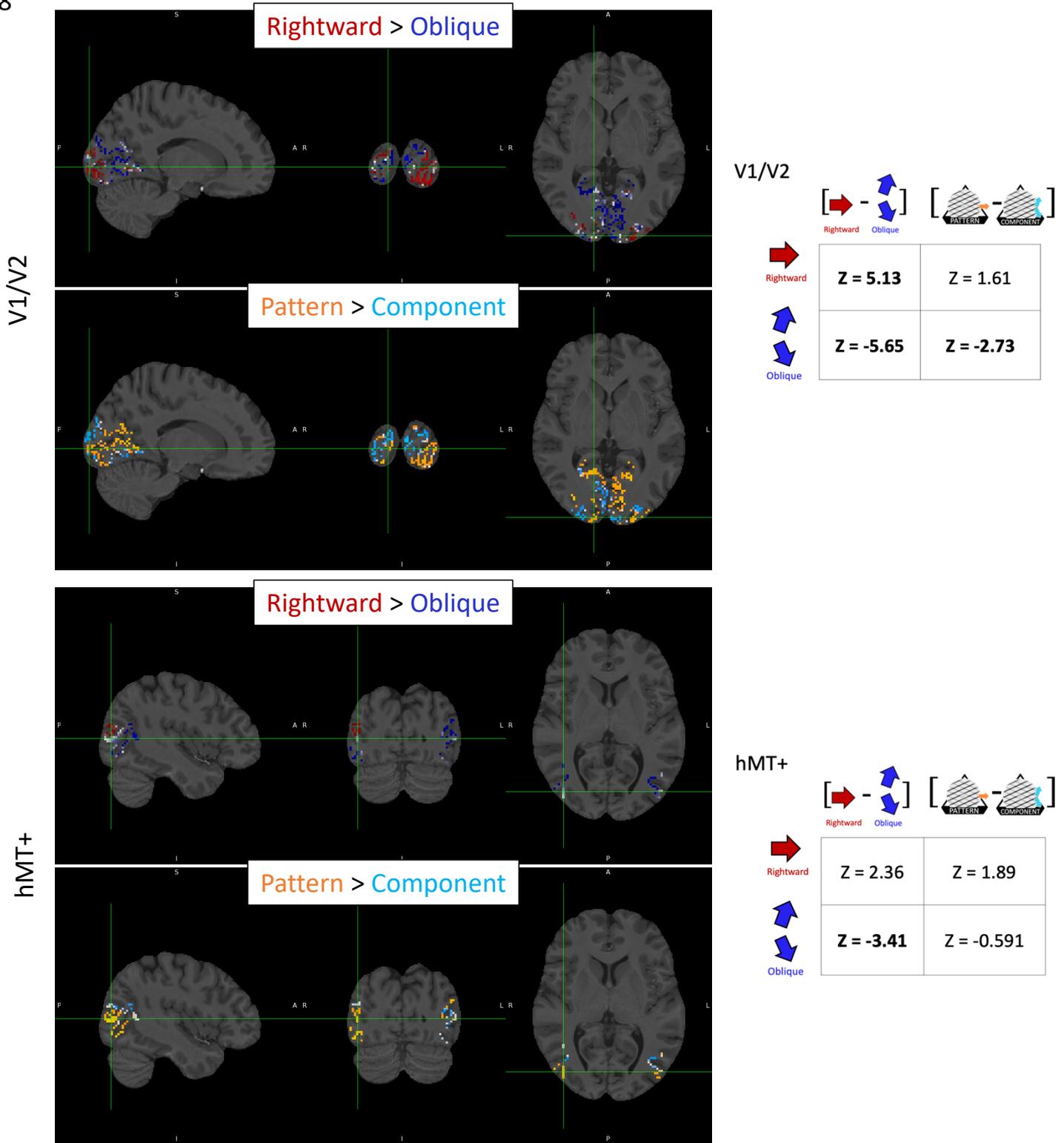


Figure 25.8: Direction-selectivity and Perceptual decision maps in subject 8
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

4.4 Group-level analyses: Linear Mixed Effect model

To account for the random effect of subject, the analyses were performed using an LME that permitted extracting conclusions at the group-level while accounting for random inter-individual differences.

We defined an LME model as follows (see diagnostic plots in Fig. 26):

$$(Y|sub) \sim \beta_0 + \beta_1 \text{Area/Subdom} + \gamma_{0,sub} + \gamma_{1,sub} \text{Area/Subdom} + \varepsilon$$

Where Y is the z-score for the Pattern vs Component GLM contrast estimate conditional on subject,

Area/Subdom is the nested fixed effect of Subdomain (Subdom: Rightward- or Oblique-selective) within each visual area (Area: V1, V2, hMT+).

The model includes both a random intercept $\gamma_{0,sub}$ for subject and a random slope $\gamma_{1,sub} \text{Area/Subdom}$ for subdomain within area. Both of these random effects are assumed to be normally distributed with mean 0 and variances estimated to best fit the model. The diagnostic plots in Fig. 26 support that residuals from the model are in reasonable accord with the model assumptions.

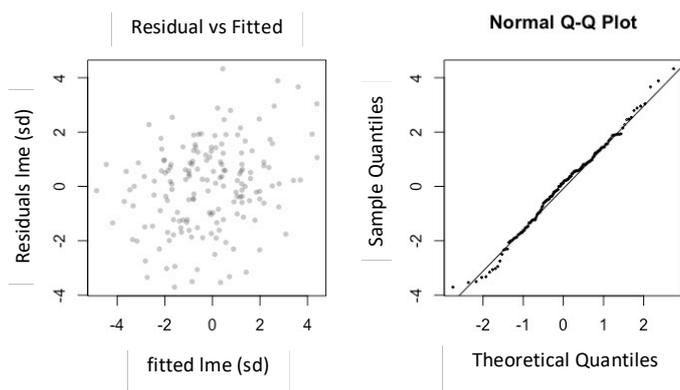


Figure 26: Diagnostic plots of LME model for Pattern/Component selectivity

The Residual vs Fitted plot shows a random distribution of residual variance around zero.

The Q-Q probability plot shows that data follow Normal distribution.

This model enabled us to test the main hypothesis that “Rightward/Oblique subdomains’ activity reflect Pattern/Component perceptual decisions” against the null hypothesis of no significant relation.

5. Results

5.1 Behavioral results

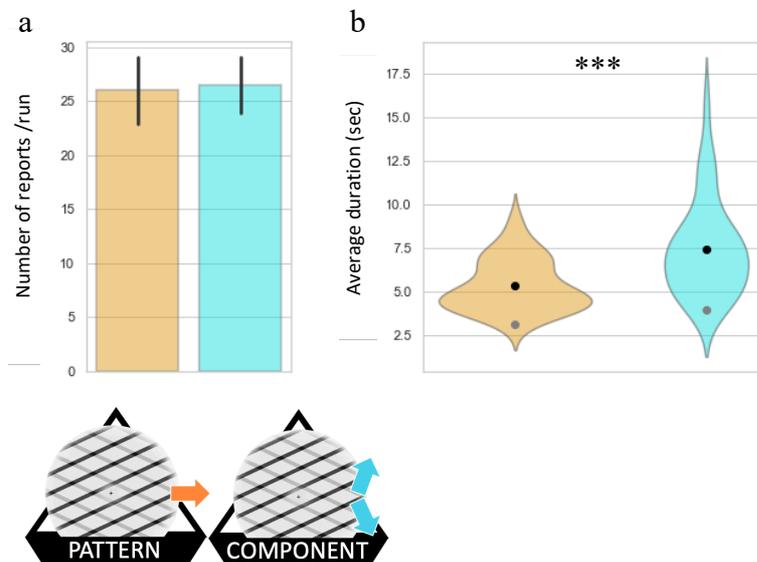


Figure 27: Duration and proportion of each perceptual states during a bistable motion stimulation

(a) Average number of occurrences per run (not significant, ns)

*(b) Distribution of the average duration of Pattern/Component percept (paired two-sample t-test $p < .001$ ***)*

N = 34 observers

Before investigating the BOLD signal, we verified that the duration and number of perceptual switches recorded from participants (Component/Pattern button presses) were approximately balanced across the experiment. Therefore, we compared the number or reported perceptual switches per run and the average duration of perceptual states across reported Component and Pattern intervals.

Figure 27a shows the average number of reports per run for component and pattern. Statistical analysis indicated no significant difference in the number of reports (26.0 ± 9.2 times for the Pattern; 26.4 ± 8.0 times for the Components; $t=0.58$, $df = 33$, $p=0.56$) (Fig. 27a). Moreover, the number of reports showed little variation across runs (as indicated by 95% confidence intervals), showing no apparent effect of perceptual learning with experience.

Figure 27b shows violin plots for the average duration of pattern and component percepts. While both percepts occurred about equally often, on average the perceptual duration was significantly longer for the Component percept (as indicated by black dots: 7.5 ± 2.7 s (mean \pm sd) against 5.4 ± 1.5 s for the Pattern; $t=4.16$, $df=33$, $p<0.001$). A comparison of the medians (as indicated by grey dots) suggests that the significance of the difference in means was possibly influenced by the greater skew of the component distributions as the more robust median values are lower for both pattern and component and more similar.

5.2 Estimate of the Pattern/Component selectivity

We aimed to test whether the BOLD responses in early visual areas (V1 and V2) and hMT+ simply responded to physical stimulus properties, or also reflected the subjective perceptual state of observers. To answer this question, we determined whether these regions exhibited differential activity depending on whether observers reported perceiving Pattern motion (in the rightward direction) or Component motion (in the oblique directions).

We estimated the Pattern versus Component selectivity z-score based on the GLM contrast estimates [Pattern > Component]. In all areas, we found a subdomain-specific modulation of perceptual state. We tested the significance of this by comparing the goodness of fit of two nested LME models, with and without the subdomain-based perceptual effect, using a nested likelihood ratio test (LRT) including all visual areas. Pattern versus Component contrast estimates significantly differed between Rightward- and Oblique- selective subdomains (LRT (all areas): $\chi^2 = 46.3$, $p < .001$) (Fig. 28). Overall, rightward-selective subdomains responded preferentially to the Pattern percept. Conversely, Oblique-selective subdomains showed a Component preference (significant in all visual areas). Interestingly, the Pattern vs Component contrast estimate was not significantly different from zero in V2 and hMT+ for the rightward domains (95% confidence intervals include 0), even though the differences between subdomains differed significantly (i.e., the interactions were significant).

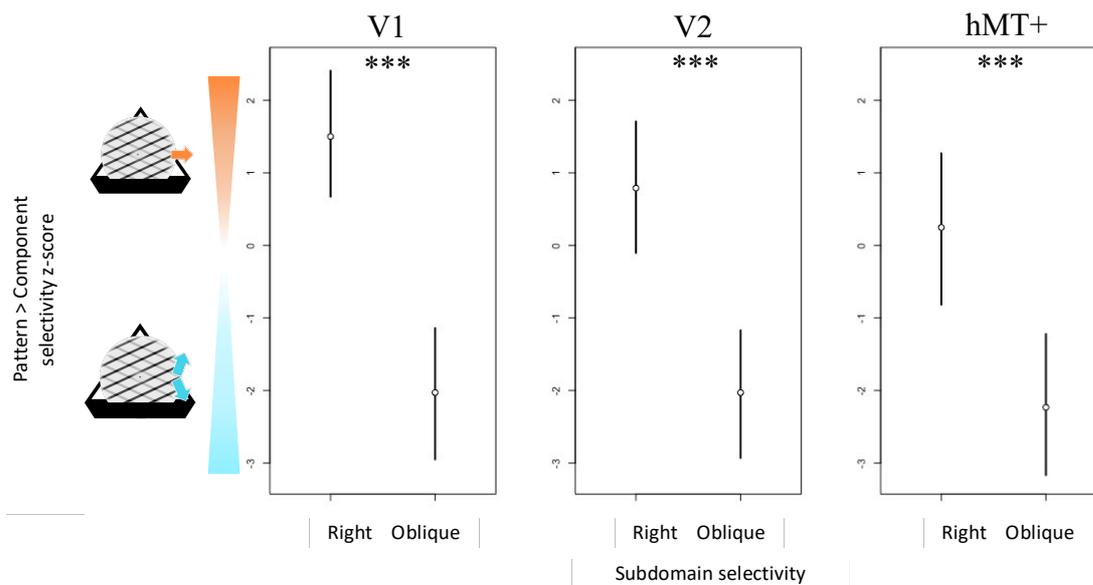


Figure 28: Pattern/Component contrast estimate within direction-selective subdomains of visual cortex
 From left to right, the score of Pattern vs Component [$P > C$] preference between Rightward (R) and Oblique (O) direction-selective subdomains in V1, V2 and hMT+ at the group-level. P vs C induced response significantly differed as a function of direction-selectivity (R vs O) in all areas.

Pattern preference ($Z > 0$) in right subdomain was significant only in V1.
 Component preference ($Z < 0$) in oblique subdomain was significant in all areas.
 Linear Mixed Effect Model comparison (LRT):
 Fixed effect of subdomain: *** $p < .001$
 N= 29 observers (5 runs)

These are marginal effects, however and, as such, are not usually interpreted in the presence of a significant interaction. Nevertheless, one can wonder if the fact that we observe a stronger selectivity in oblique subdomains consistently across areas suggests that the neural mechanism underlying switching percepts operates differentially in relation to the orientation mechanisms sensitive to the components (here the oblique directions) rather than the pattern sensitive direction (here represented by the rightward subdomains).

6. Conclusion

In this Chapter, two fundamental aspects of bistable motion perception were presented. Behaviorally, we found that for the conditions that we defined in the previous chapter, the temporal distribution of Pattern/Component perceptual states was balanced. This constituted the critical criterion in deciding to include or not participants in the statistical behavioral and fMRI analyses. Unbalanced bistable perception (i.e., an experimental condition with fewer repetitions in time) could have generated a bias in the estimation of Pattern vs Component selectivity, that is the main result of interest.

For the functional neuroimaging, it was possible to localize and optimize the size of direction-selective subdomains within visual areas of most observers. A few observers for whom this was not possible were excluded from group-level analyses. At the group-level, we found a significant interaction between the activity in the direction-tuned subdomains and perceptual state. The Pattern vs Component contrast in rightward subdomains was significantly higher than that in oblique subdomains, thus, indicating that *perceived* direction was associated with the activity of specific directional subdomains. The key finding was that this effect was found as early as area V1, with oblique subdomains preferring Component percept while rightward subdomain was tuned to Pattern percept. Area V1 is presumed to not contain pattern-selective motion detectors, therefore we hypothesize that this result reflects the existence of motion integration signals being fed back from higher-level visual areas (Fig. 28). In V2 and hMT+, we consistently found the preference of Oblique subdomains for Component percept, and the significant difference between Oblique and Rightward subdomains' [Pattern > Component] contrast estimates. However, we could not replicate the preference of Rightward subdomain for the Pattern percept (contrast estimate z-value was not significantly superior to zero). One speculation is that the neural units responding to Component motions might be more influential than those responding to the Pattern in driving the perceptual switching. This is consistent with the results of our previous classification analyses within motion-selective visual regions suggesting that the neural mechanisms underlying bistable motion perception are specific and distinct from responses of direction-tuned units to unambiguous Rightward and Oblique moving stimuli.

The general results will be more thoroughly discussed in the general conclusion. In the next chapter, additional analyses were implemented to examine potential influences of eye movements on the results.

CHAPTER 4:
**Analysis of a potential role of Eye Movements
in Bistable Motion Perception**

1. Effect of eye movements in bistable perception

Motivation

Several studies support the theory that local competition between multiple mental representations of a stimulus is the driving force for bistable perception, as an attempt to interpret noisy sensory signals based on statistical regularities (see comment by Hupé et al., 2008). Nevertheless, alternative hypotheses formulate a role for involuntary eye movements in triggering perceptual switches. Some research studies investigating the neural correlates of bistable perception have shown associated oculomotor physiological markers that predicted perceptual state (Wilbertz et al., 2018; Brych et al., 2021).

1.1 From sensing to perceiving

Filling in perceptual gaps

Primates produce a variety of eye movements aiming to explore the surrounding visual environment and to build internal representations. This is achieved through both active and passive sensing. In terms of eye movements, active scanning refers to voluntary oculomotor activity that is intended to explore the environment either globally or by selectively focusing attention on local features. Passive scanning, however, consists in small reflexive eye movements automatically induced in response to visual stimulation. Perception arises from the extraction of the most salient/relevant information based on both external and internal factors. Due to the fact that visual scanning is punctuated by eye movements, blinks and occluded visual targets, the sensory input contains gaps (i.e., periods where the sensory signal is either partially or completely blocked, distorted or attenuated). Consequently, re-creating a smooth perception from the sampled information requires computational processes of perceptual inference and interpolation. It was proposed that this perceptual filling-in is made possible by inhibitory signals, originating from subcortical oculomotor regions (Reppas et al., 2002; Ramcharan et al., 2001; Berman and Wurtz, 2011; Frost & Niemeier, 2015), that would suppress motion sensitivity specifically in hMT+ complex (and other motion-selective dorsal areas) around the time of the gap (Ibbotson et al., 2008; Bremmer et al., 2009; Berman and Wurtz, 2011). Following the idea that the world is filled with statistical regularities, and assuming that our brain machinery makes predictions based on prior information, it seems reasonable to assume that it is unlikely that a major change will occur during the sensory gap (von Helmholtz, 1967). In principle, highly predictable stimuli (e.g., stationary objects, repeated items, ordered sequences, etc.) will be easy to fill-in, e.g., by interpolation, extrapolation or repetition of information at the gap boundaries. Nonetheless, the underlying neurocomputational processing becomes more complex when dealing with unexpected events.

Visual tracking of moving targets

When tracking a moving stimulus and depending on the nature of the signal (variation of speed, size, direction, contrast, etc.), the oculomotor system adapts to the object's motion, enabling the formation and updating of the image of the target on the retina (for reviews see Krauzlis, 2004; 2005). Different types of eye activity are elicited depending on the speed of the target. When the target moves slowly (velocity < 30°/s), smooth pursuit eye movements are made to track the object's motion while stabilizing the image at the fovea (i.e., region of the retina with

the highest visual acuity). However, for higher speeds, the tracking may not keep up with the target, and the need to “catch-up” leads to rapid ballistic eye movements, called saccades, towards the object.

Tracking through smooth pursuit

We have seen that eye velocity is closely related to retinal motion signals as it adapts to the target motion. Smooth eye pursuit corresponds to a continuous activity characterized by slow rotations (eye velocity < 50 deg/s) of the eyes made to compensate for a target’s (slow) motion.

In Humans, smooth pursuit eye movements in response to a moving visual target usually have a latency of about 80–120 ms (Carl and Gellman, 1987, Krauzlis, 2004), although it can vary as a function of the stimulus properties (e.g., Tychsen & Lisberger, 1986; Ferrera and Lisberger, 1995; Krauzlis et al., 1999; Spering and Gegenfurtner, 2008) and be influenced by predictions of the target trajectory (e.g., Bahill & McDonald, 1983).

Although it is continuous, smooth pursuit is composed of an initial phase (~100 ms) during which eye velocity is driven by the retinal image before it adapts to the velocity of the target. Concretely, an accelerated movement is first made towards the target, and then, the object of interest is maintained and stabilized at the fovea (Lisberger et al., 1987, Tychsen and Lisberger, 1986).

“Catching up” with saccades

For higher-speed motions, correctional or “catch-up” saccades are made to compensate for retinal image slippage (De Brouwer et al., 2002). The visual system is equipped to track motion signals up to a velocity of 100°/s (Meyer et al., 1985). Unlike smooth pursuit, saccadic eye movements refer to discrete, ballistic movements that enable fast displacement towards the target of interest.

Pupil dilation

Pupillometry, or the measurement of changes in pupil diameter as controlled by the autonomic sympathetic nervous system, is a relatively old method that has been shown to reveal valuable information on neural activity. Indeed, pupil dilation in the absence of luminance changes is a physiological marker that has been associated with attention, interest, emotion, or more generally with mental workload and arousal (Iqbal et al., 2004; Tullis & Albert, 2008).

Changes in pupil size have been associated with specific visual processing such as the transition from component to pattern motion and vice versa (Sahraie & Barbur, 1997). Interestingly, this phenomenon persists even in the case of internally-driven perceptual alternation, while visual stimulation remains unchanged. Conducting tasks of bistable motion perception, Einhäuser et al. (2008) reported an increase in pupil diameter, mediated by a release of norepinephrine (NE) in the locus coeruleus (a subcortical structure), that coincided with the time of perceptual switch. Notably, they suggested that pupil dilation was a reliable predictor of the stability (i.e., duration) of the upcoming perceptual state. Two more recent studies that aimed to replicate these results following the same methodology (Lamirel et al., 2008; Hupé et al., 2009) found

the same variations of pupil dilation around the time of perceptual switching, but it was argued that the predictive effect of pupil response was mainly an artefact of very short percepts (< 3 sec), where variations of pupil diameter overlapped from one percept to the next. It resulted that discarding these short percepts fully abolished the positive correlation. Alternatively, they proposed that motor responses (button presses) and blinks explained most of the pupil modulation over time.

1.2 Eye control and perception in visual cortex

A brief review

It is commonly accepted that sensory signals ascending and descending the visual cortical hierarchy participate in progressive levels of information processing over different spatial scales and complexity levels. Therefore, activity recorded within visual cortex can reflect sensory input, higher-level mental representations, but also attentional and oculomotor activity. I will not elaborate on attentional control in this chapter, but Wang & Kowler (2021) demonstrated that smooth pursuit activity is a reliable feature in estimating the distribution of attention during perceptual tasks involving dynamic display.

Bogadhi and collaborators (2011) proposed a Bayesian framework explaining oculomotor activity in the light of two recurrent cortical loops involved in motion integration, with area hMT+ playing a central role. Such visual processing illustrates two important properties of oculomotor function. First, while retinal motion information is being collected through smooth eye pursuit, fast corrections (saccades) of the eye tracking error /delay can occur (possibly through hMT+-to-V1 feedback) (Bayerl & Neumann, 2004; Montagnini et al, 2007). Second, it was described that the inhibition of visual input signals to smooth out transient changes in the image (i.e., perceptual gaps) utilizing higher-level non-retinal signals originated from higher-level cortical areas (possibly through Frontal Eye Field-to-hMT+ feedback) (Newsome & Paré, 1988; Ilg & Their, 2003).

Because they are intricately entwined, disentangling neural correlates of perception from eye movement-related activity remains challenging. However, it can be a confounding factor when investigating FB function. Interestingly, electrophysiological studies in awake monkeys revealed systematic differences in the temporal profile of activity in hMT+ neurons: while motion integration signals (representing pattern motion) were recorded within 60 ms post-stimulus, smooth pursuit activity occurred after 100 ms approximately (Pack & Born, 2001). This finding suggests that the oculomotor response may have been a result of a change in perception (e.g., from component to pattern percept) instead of the reverse. This observation was subsequently confirmed by other research studies that experimentally investigated eye movements independently of conscious perception. An example of this is that small fluctuations in the input signal are not always consciously perceived even though they are processed by the oculomotor system, as indicated by triggering of eye movements (Tavassoli & Ringach, 2010).

It remains undetermined if oculomotor activity and motion processing depend on similar mechanisms or are processed in parallel (Stone and Krauzlis, 2003 ; Dürsteler and Wurtz, 1988 ; Komatsu and Wurtz, 1988; 1989 ; Ilg and Thier, 2003). This question cannot be resolved by fMRI as this technique does not provide a sufficient temporal resolution to accurately discriminate between the two signals. However, a recent study conducted a series of experiment

aiming to determine the causal effects of ocular following response to moving gratings through apertures of varying size, orientation and/or direction. They demonstrated that directional eye movements were affected by orientation-selective mechanisms rather than direction (Sheliga et al., 2015), in a way that reflects the end-stopping property of complex cells in V1 (Cavanaugh et al., 2002). Therefore, we expected that the oculomotor activity during a task of bistable motion where only direction (and not orientation) is perceived differently over time would not constitute a perceptual bias. To ensure constant visual stimulation during the viewing of the bistable moving plaid, we limited eye movements (via a fixation task) and controlled for involuntary residual eye activity. This enabled to investigate what if any role eye movements may play in the perceptual switching in bistable motion perception.

As discussed in the introduction when performing motion integration, eye movements tend to follow the dominant perceived direction. Initially, they are driven by local component motions, hence conforming to the Vector Sum model prediction (Chen et al., 2005). After approximately 200 ms, eye tracking progressively adjusts to the object's global velocity vector which yields ocular response closer to the IOC prediction. However, these conclusions resulted from experiments with short presentation times and allowing voluntary eye movements (Wilbertz et al., 2018) (i.e., there was no fixation task), therefore, such an effect, if it exists, is expected to be weaker when performing a fixation task.

Isolating perceptual decisions from oculomotor activity

Isolating the activity reflecting perceptual decisions from oculomotor responses is challenging if not impossible. However, in order to maximize the chance, we designed an experiment with a highly controlled eye tracking environment. To achieve this, we gave our participants a somewhat simple task: they had to lock their gaze on a cross in the center of the screen (hence avoiding voluntary eye movements) while reporting their perceptual switches. However, involuntary oculomotor response remained and should be accounted for. Based on the analyses of the eye movements in relation to perceptual state, we tested for any significant bias in the eye movements made within a particular state, or just before a perceptual switch. Finding a significant effect of the eye movements would impose to add this explanatory variable to the list of confounding factors in our fMRI model design matrices.

Investigating involuntary eye movements

In contrast to voluntary pursuit and saccadic eye movements, involuntary eye movements refer to reflex oculomotor activity and occur at shorter latencies and over smaller amplitudes (Miles et al., 1986). Nonetheless, they are driven by similar neural mechanisms as it was reviewed extensively elsewhere (e.g., Ilg, 1997; 2002, Keller and Heinen, 1991, Krauzlis, 2004; 2005, Thier and Ilg, 2005).

In theory, when asked to fixate a stationary target stimulus during the presentation of a moving stimulus, observers who perform well at the task will succeed in inhibiting their voluntary oculomotor responses; however, they will have no control over involuntary residual movements. One example of such reflex movement is Optokinetic Nystagmus (OKN) that is characterized by two phases: a slow eye drift that consists of a smooth movement made in the direction of a moving target, and a fast correctional (micro)saccade back to the fixation target (see example in Fig. 29).

OKN has been shown to provide a reliable estimate of perceptual state in bistable perception (Enoksson, 1963 ; Watanabe, 1999 ; Wei & Sun, 1998 ; Naber et al., 2011). Thus, it behooves us to examine its relation to perceptual state in bistable motion.

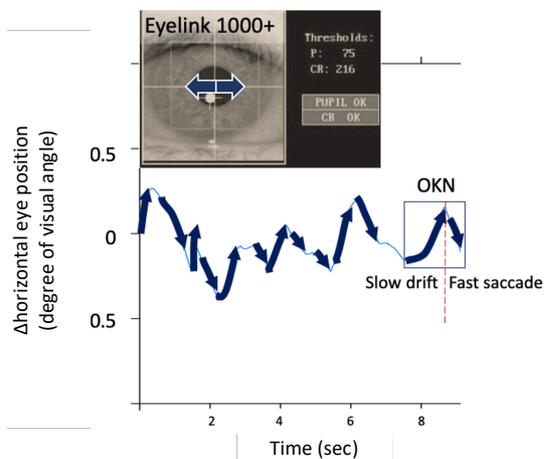


Figure 29: Extracting eye movements from the eyetracker recording. The plot shows an extract of the left eye's horizontal displacement (in deg of visual angle) over time recorded during a session of bistable motion perception.

Upward and Downward arrows indicate rightward and leftward eye movements, respectively. The trace illustrates the two phases of OptoKinetic Nystagmus, that consist of a slow drift in the direction of the moving target, followed by a fast saccade back to fixation.

2. Material & Method

Data collection

We analyzed the eye movement data of all the individuals who were included in the fMRI group-level analyses, excluding 5 participants whose eyetracker recordings experienced technical problems, resulting in a dataset of N=24 observers. An Eyelink 1000+ setup synchronized to the fMRI scanner tracked oculomotor activity (raw eye position, saccades, blinks, pupil dilation). The Eyetracker system was set on monocular mode (tracking left eye), and a 9-dot calibration was performed before data collection began, to ensure high gaze accuracy and quality of recording. Eyetracking constituted an important measure to obtain during the whole session, as it has been shown that in the absence of an attentional/fixation task, eye movements reliably predict perceptual state (Wilbertz et al., 2018).

Extraction of eye-velocity features

In spite of the fact that observers were trained to fixate on a central target, involuntary eye movements persisted in response to the moving stimulus. We extracted both the speed and gaze direction following a methodology that was inspired by Wilbertz and collaborators. We extracted two-dimensional raw eye position (pixels) per millisecond, that was converted to degrees of visual angle (deg) based on distance to the screen and pixel size. In order to realign the eye data timeline to behavior (button presses), we added a 500-millisecond time lag to the eyetracker recordings. Then, we excluded blink periods as assessed by the Eyelink 1000+ system (\pm margin of 10 additional ms before and after). The velocity was computed as the difference between two successive eye positions divided by Δ time (1ms). We excluded artefacts that were mostly due to technical signal loss, and selected involuntary eye movements by removing any data where the rolling mean of absolute velocity over a 500-millisecond window was $< 15 \text{ deg.s}^{-1}$ and the absolute acceleration $< 100 \text{ deg.s}^{-2}$. Subsequently, we applied

a smoothing filter to the selected eye velocity by computing rolling means over 500-millisecond window. We repeated this process for both vertical and horizontal dimensions.

3. Analyses pipeline

3.1 Analyses 1: Predicting perceptual state from previous eye movements

In the context of the bistable moving plaid, observers experienced one of two percepts, either the Pattern (0° , rightward) or the Components ($\pm 62.5^\circ$). To exclude the possibility that involuntary eye movements accounted for the neural correlates of the perceptual decisions, we investigated horizontal eye movement velocity over the 1 second interval before a perceptual switch and compared Pattern and Components conditions.

We fitted a model representing perceptual state as a function of eye velocity and performed decoding analyses in which we estimated how accurate model predictions were in classifying Pattern from Component perceptual state. We hypothesized that if there was an effect of eye movements relative to the perceptual state, this would be reflected by a specific pattern in eye movements made just before perceptual switches (Wilbertz et al., 2018).

For each subject, we extracted the average standard deviation of the eye movement velocity over the 1-second interval before each perceptual switch along both horizontal and vertical

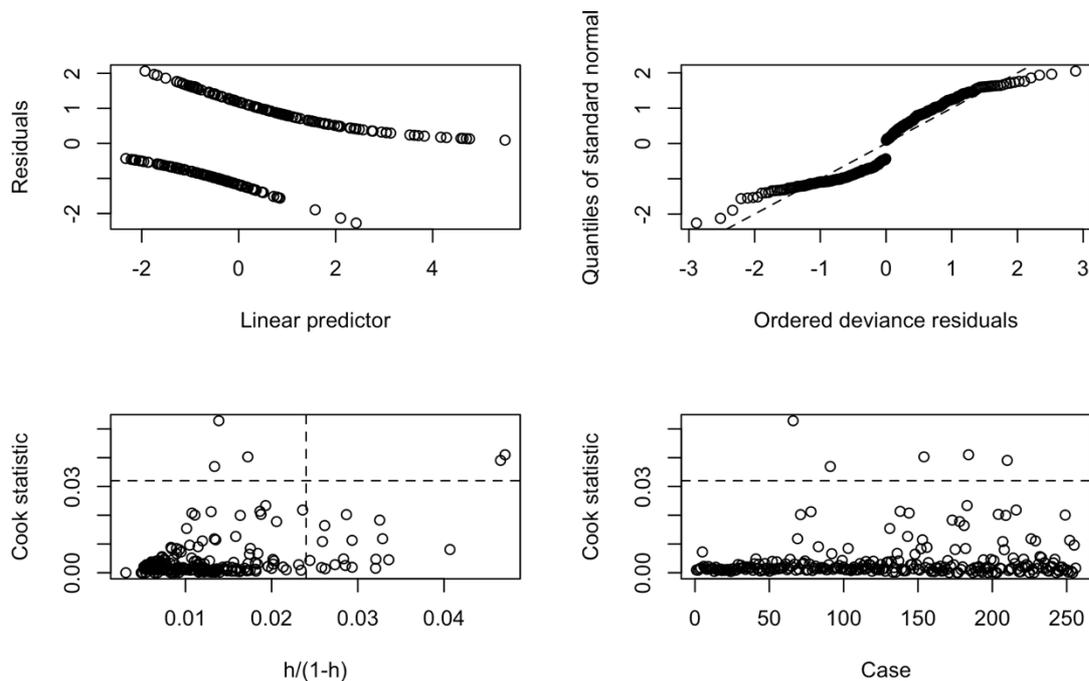


Figure 30: Diagnostic plots of binomial glm representing perceptual state as a function of velocity features
 Linear predictor (aka "fitted") vs residuals (using jackknife deviance residuals, computed as $\text{sign}(\text{dev}) * \sqrt{(\text{dev}^2 + h * \text{rp}^2)}$ where dev is the deviance residual; h is the hat value; and rp is the standardized Pearson residual).

Ordered deviance residuals vs quantiles of standard normal: "Q-Q plot" (using standardized deviance residuals)
 Leverage ($h/(1-h)$) vs Cook statistic, a more conservative rule to decide which influential points might be
 Outliers Case vs Cook statistic.

dimensions (σ_x and σ_y , deg/s). We fitted a binomial glm with a logit link estimating perceptual state (*Resp*, P or C) as a function of the x and y eye velocity' standard deviations for the group. Diagnostic plots show that fits were not affected by influential outliers or collinearity between explanatory variables. The unusual form of the diagnostic plots is typical for binomial glm models, which are interpreted based on Gaussian models of error rather than the binomial deviance residuals used in the plots in Fig 30.

It was found that σ_y and σ_x significantly differed across conditions showing bigger velocity variations for Pattern than for Components.

The classification model was fitted to the data and estimated coefficients are as follows:

Model: $Resp \sim \beta_0 + \beta_1\sigma_y + \beta_2\sigma_x + \varepsilon$, family = binomial)

Coefficients:

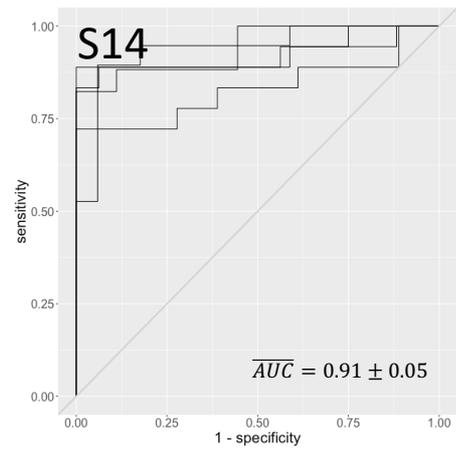
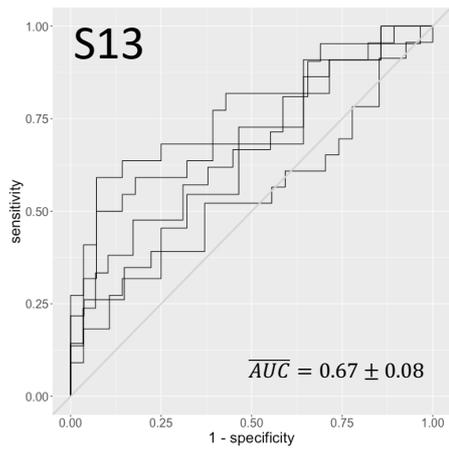
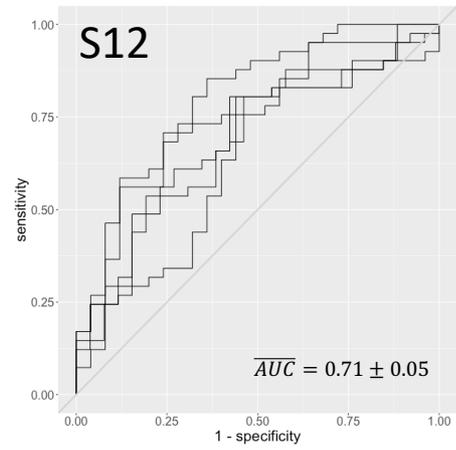
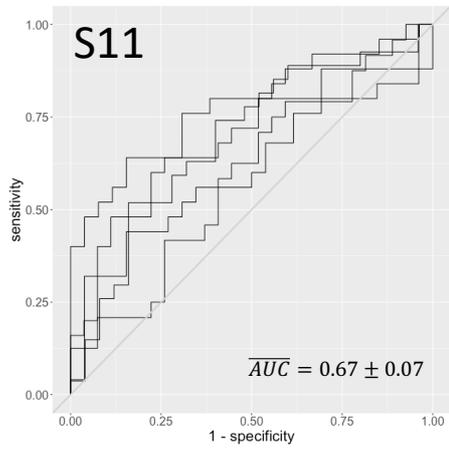
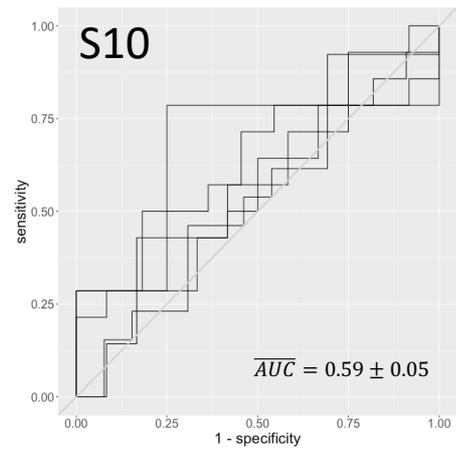
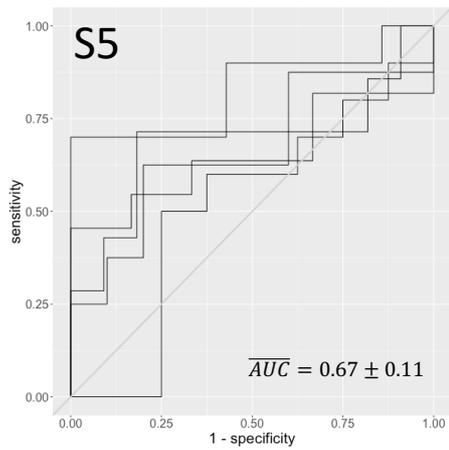
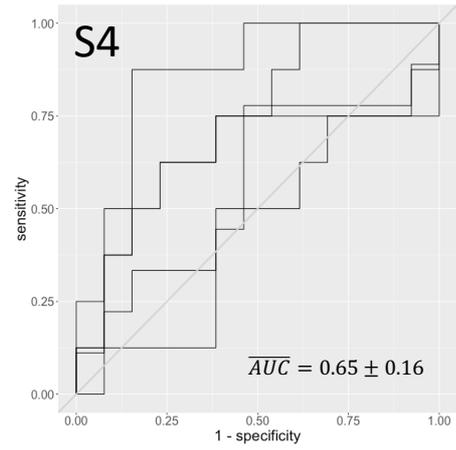
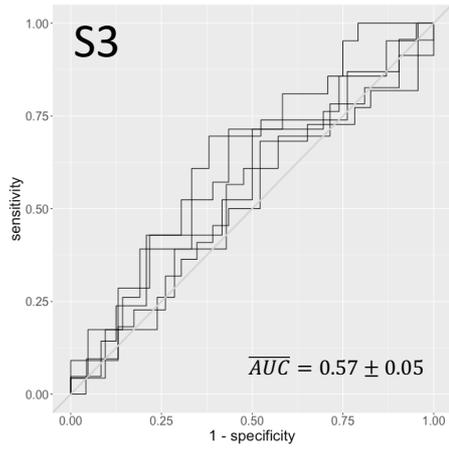
	Estimate	SE	z-value	Pr(> z)
β_0	0.1013	0.1582	0.640	0.52
σ_y	0.6166	0.1028	5.998	2e-09 ***
σ_x	0.5255	0.1519	3.461	5.4e-04 ***

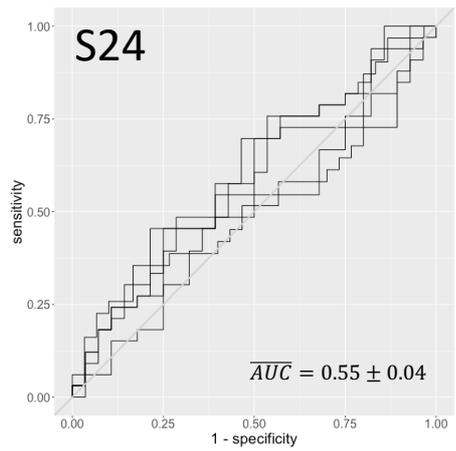
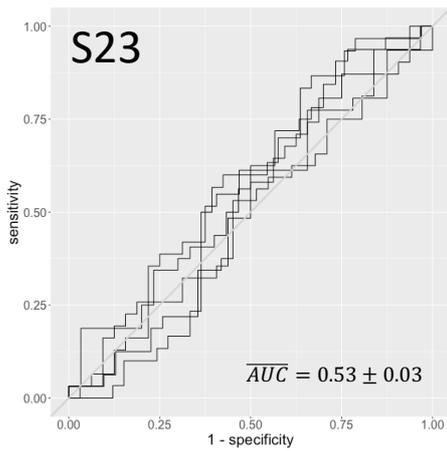
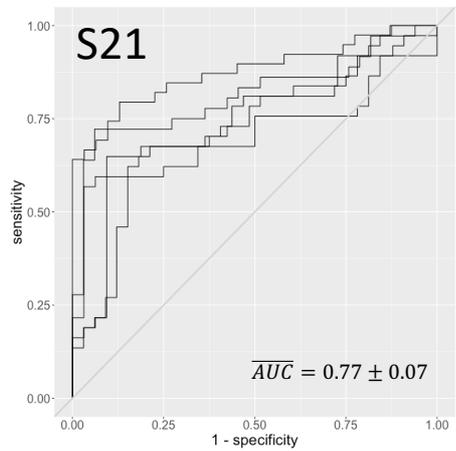
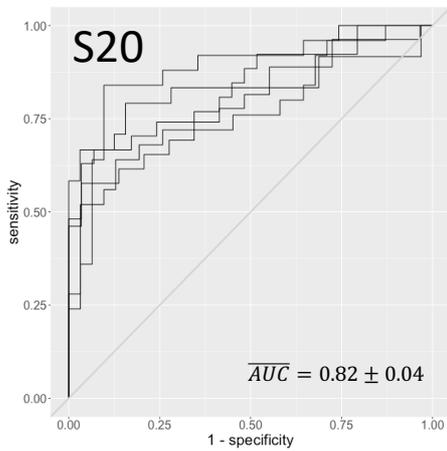
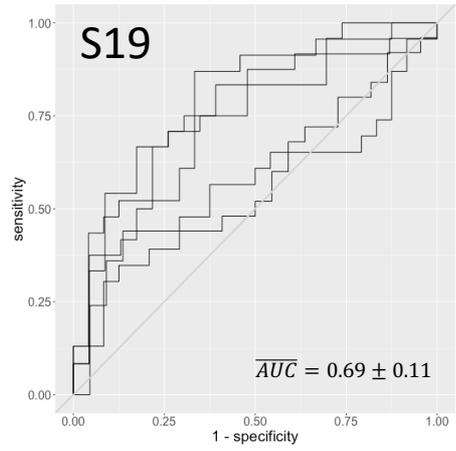
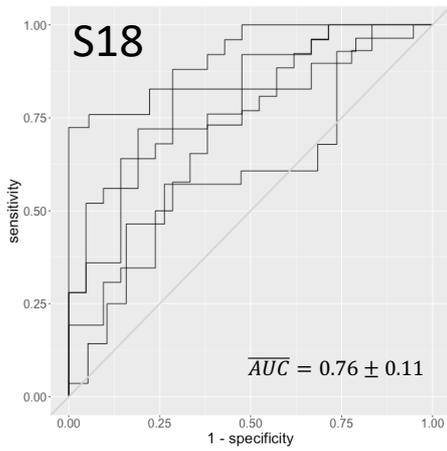
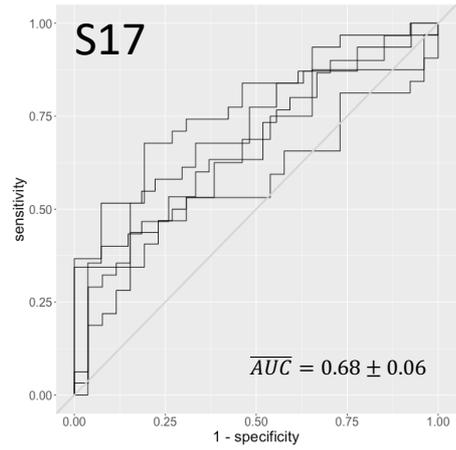
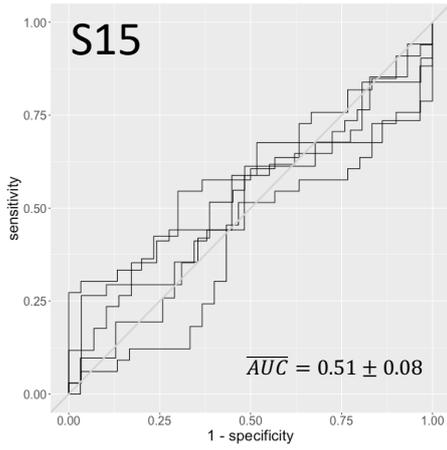
Significance code: *** p < 0.001

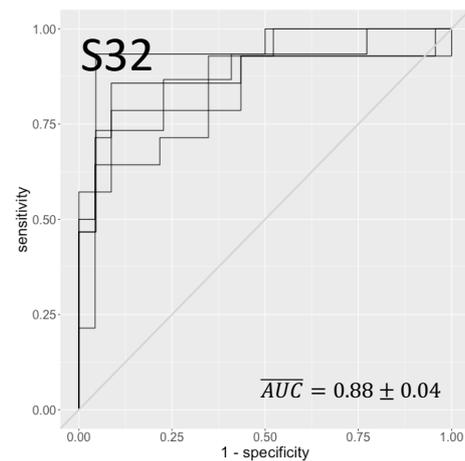
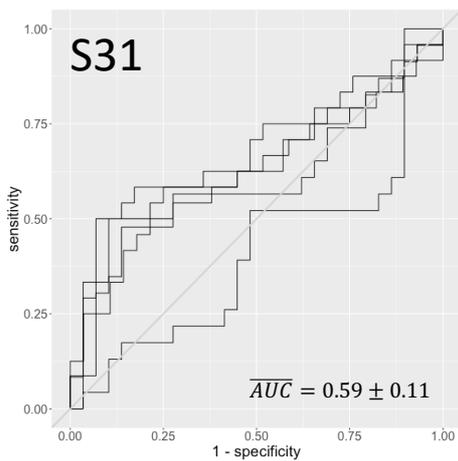
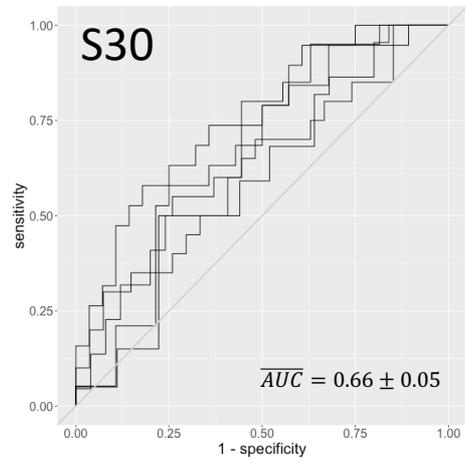
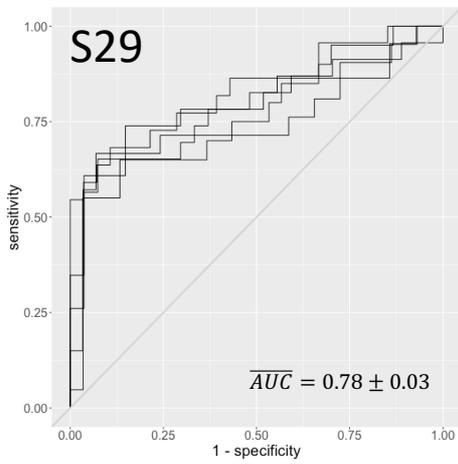
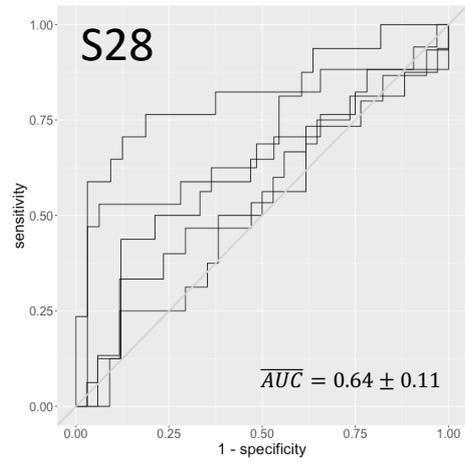
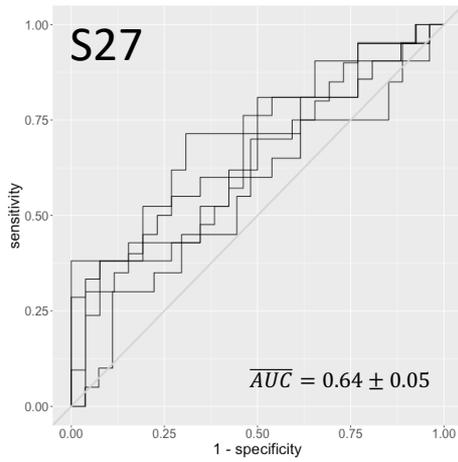
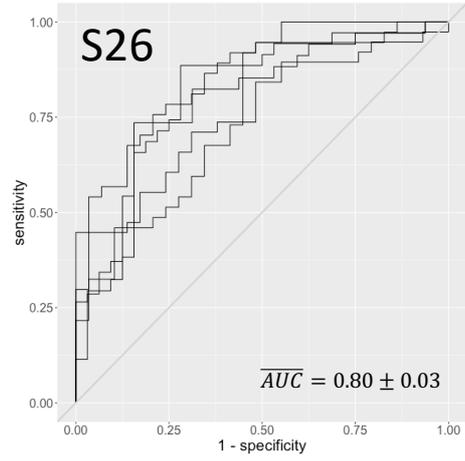
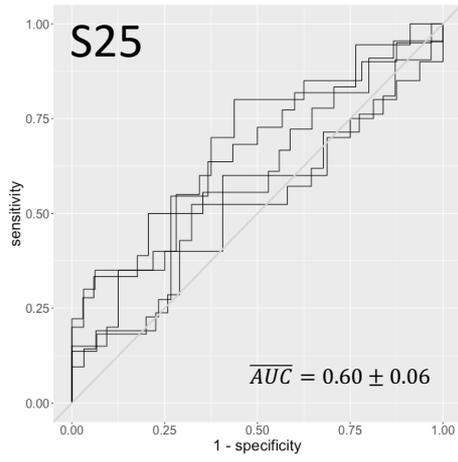
With *Resp* representing perceptual state as reported by observers,
 β_0 , the intercept (*Resp* = Component)
 σ_x and σ_y , the linear predictor variables,
 β_1 and β_2 , regression coefficients
 ε , Gaussian error accounting for the residual variance

We chose to perform within-subject decoding analyses to account for individual heterogeneity. As both σ_x and σ_y were correlated to perceptual state, we used these two explanatory variables to classify the response variable, perceptual state (Pattern/Component). We conducted a 5-fold cross validation decoding analysis (leave-one-run-out), training the model on Pattern vs Component discrimination based on the two velocity-based features

Below (Fig. 31), is represented the 5-fold cross validation ROC curves for each observer, with the mean and standard deviation of the classification scores. The light grey line represents chance-level performance (AUC score = 50%).







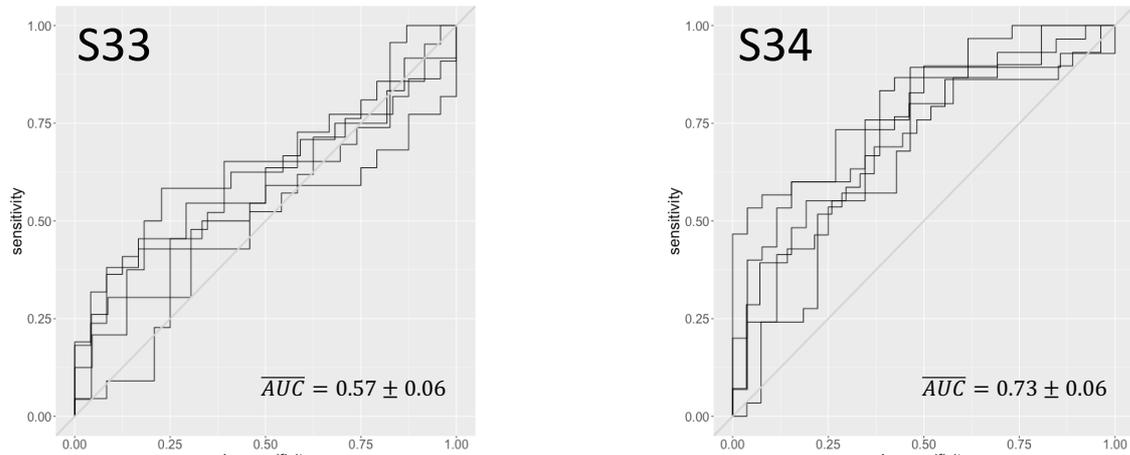


Figure 31: Subject-level classification analyses of perceptual state as a function of pre-switch eye velocity. The average 5-fold AUC score, that is the area under the ROC curve (sensitivity \sim (1- specificity)) reached significance ($AUC \pm sd > 50\%$) for most observers.

This indicates that pre-switch eye velocity features performed better than chance at predicting the perceived motion (P or C).

The average AUC score for the group was 68% with a 95% confidence interval $CI_{95} = [61\%, 75\%]$, falling into the category of “poor-to-acceptable discrimination”.

However, because classification was significantly superior to chance-level in most subjects, we need to compare the eye movements recorded during the bistable perception to a task that showed very little eye movement dependence: the localizer task. Indeed, because the localizer task consisted of short presentations of single gratings moving unambiguously, observers performed particularly well at fixating the cross in the center of the screen. However, due to the fact that there was no perceptual switch during the localizer sequence, the comparison was made on the basis of the eye movement variations occurring within (not before) a particular condition (Localizer: Right vs Oblique trials; Bistable: Pattern vs Component percept).

3.2 Analysis 2: Comparison Localizer/Bistable perception

In order to compare between tasks, we extracted the same variables as previously, σ_x and σ_y . We hypothesized that any dominance in vertical/horizontal eye velocity variation would be significantly correlated to the perceptual state. Due to the fact that the Component percept (or Oblique condition for the localizer) contains two concurrent directions that are symmetrical in relation to the x axis, $\bar{\sigma}_y$ was expected to be higher than $\bar{\sigma}_x$. Conversely, during the Pattern percept (main task), as only rightward motion is experienced, we expected the reverse trend, with $\bar{\sigma}_x$ being predominant. These are the expected results if there was a bias in the direction of the eye movements during bistable perception. Addressing our initial question, we tested the correlation between vertical/horizontal eye velocity variations and perceived motion for each observer (relative to Pattern/Component percepts or Rightward/Oblique trials).

We used Deming regression to explore the relation between the two spatial dimensions (x,y) of the eye movement dataset. This is implemented in the R package deming using the deming function (<https://cran.r-project.org/web/packages/deming/index.html>).

Deming regression, also called total and principal components regression, is a method aiming to estimate the line that best fits a dataset when both variables are subject to measurement error, as is the case here.

The slope, $\hat{\beta}_1$ was estimated based on the following equation:

$$\hat{\beta}_1 = \frac{s_{\sigma_y} - \delta s_{\sigma_x} + \sqrt{s_{\sigma_y} - \delta s_{\sigma_x} + 4\delta^2 s_{\sigma_x}^2 s_{\sigma_y}}}{2s_{\sigma_x} s_{\sigma_y}}$$

With s_{σ_y} , the estimated variance of σ_y ,

s_{σ_x} , the estimated variance of σ_x ,

$s_{\sigma_x \sigma_y}$, the covariance of σ_x , and σ_y ,

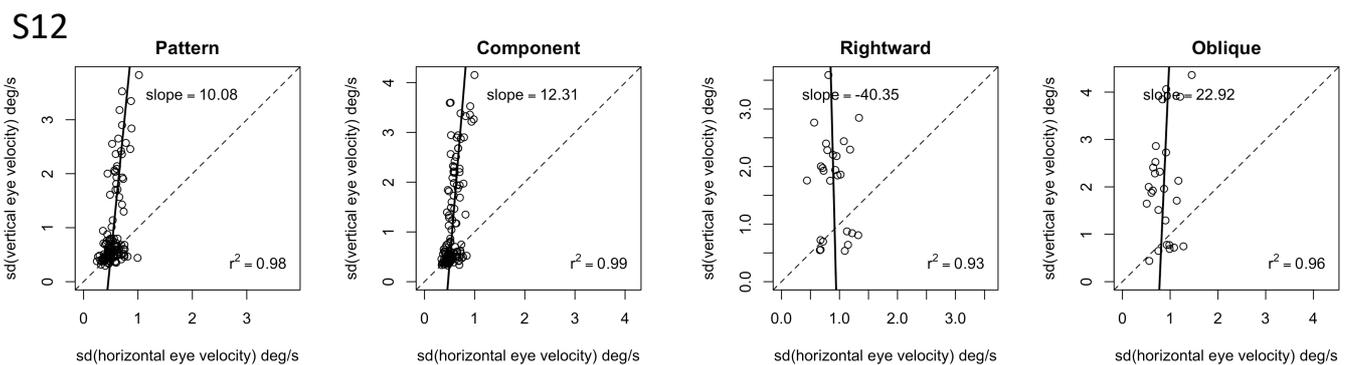
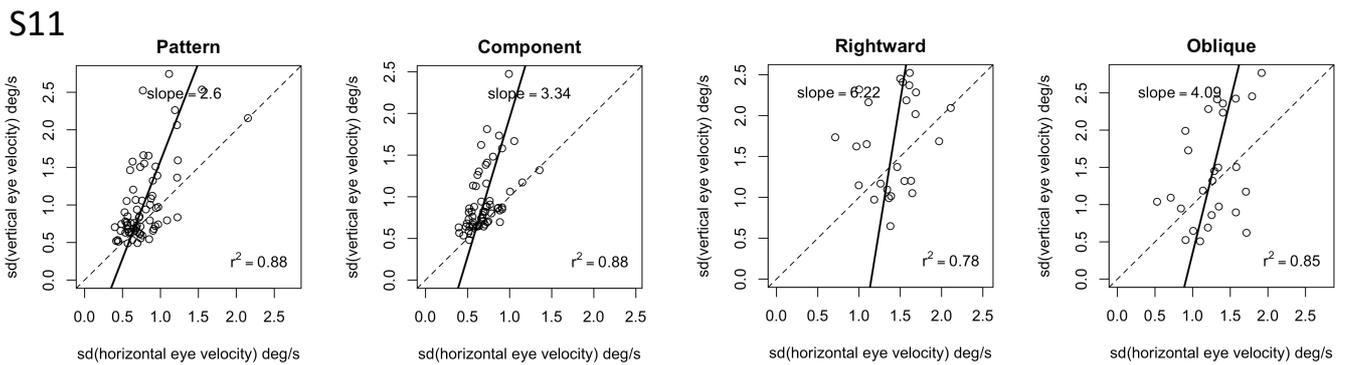
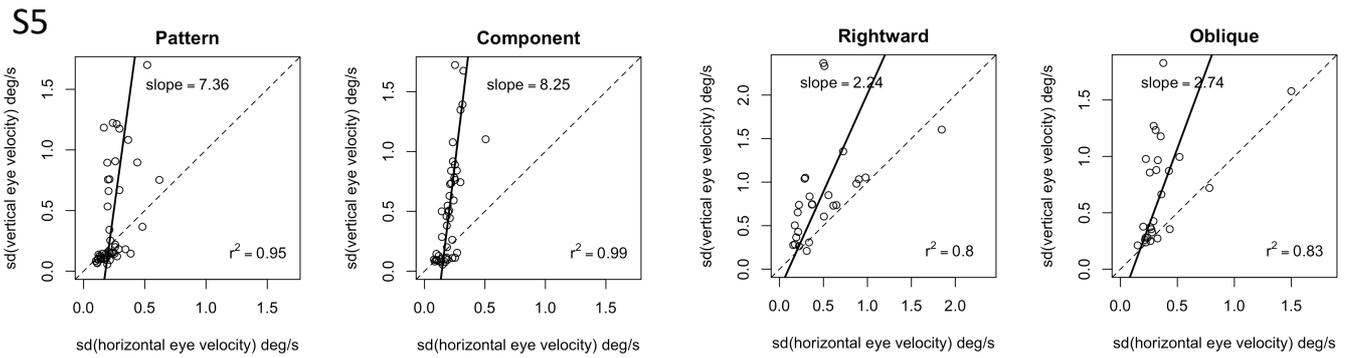
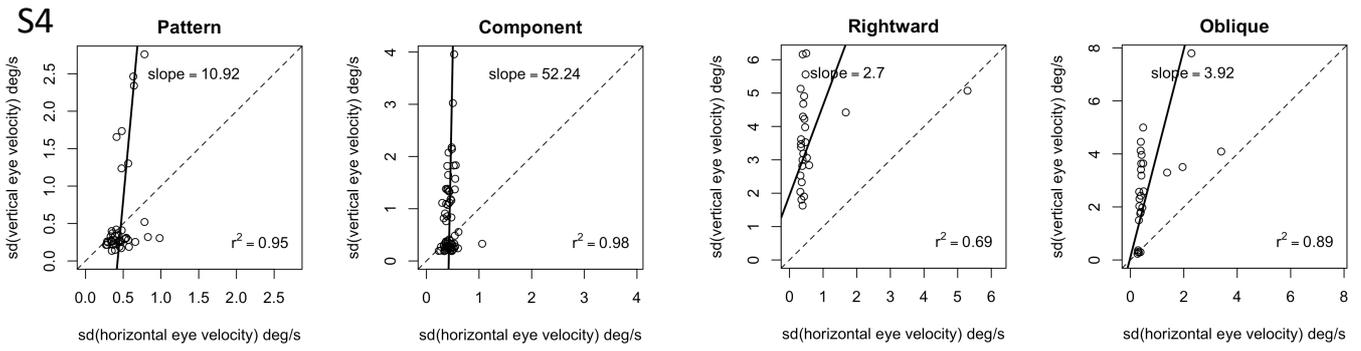
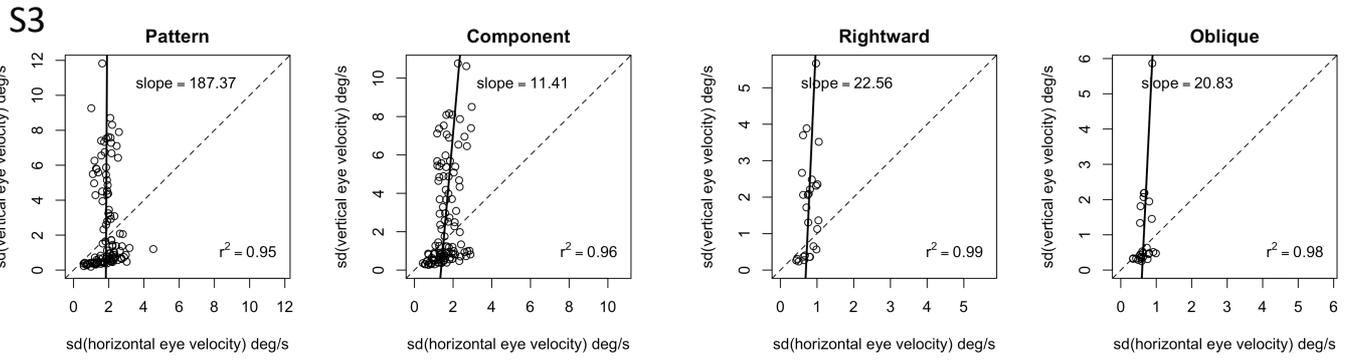
δ , an error term ($\delta = 1$ as measurement methods were equivalent)

The intercept $\hat{\beta}_0$ was defined as:

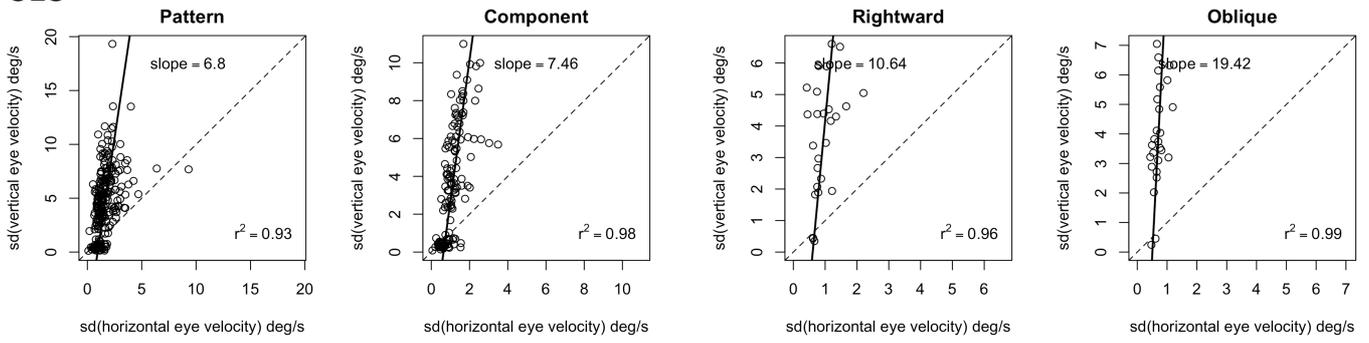
$$\hat{\beta}_0 = \bar{\sigma}_y - \hat{\beta}_1 \bar{\sigma}_x$$

In general, most observers showed stronger variations of vertical eye velocity, no matter which motion direction was perceived. In other words, eye velocity variations were not relevant to perceived motion. At first glance, it seems surprising that vertical eye movements dominated over horizontal ones, even when horizontal motion was perceived. One possible explanation is that that eye movements were made in the direction of the Fourier energy of the component gratings, hence reflecting first-order physical properties of the stimulus rather than perceived motion direction. This would be the case if the eye movements were responding to units with small oriented receptive fields subject to an aperture effect (Sheliga et al., 2015).

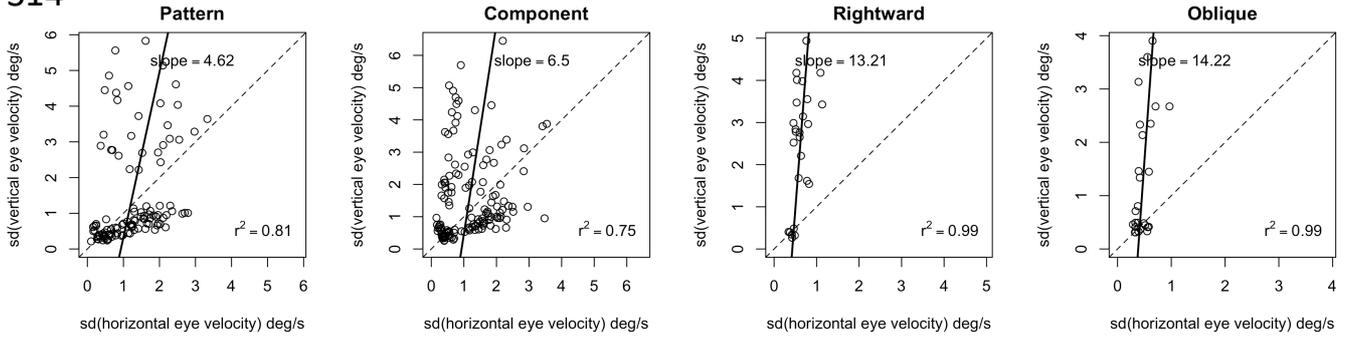
The individual Deming regression profiles are illustrated in Fig. 32 for the bistable perception (columns 1-2) and the localizer task (columns 3-4).



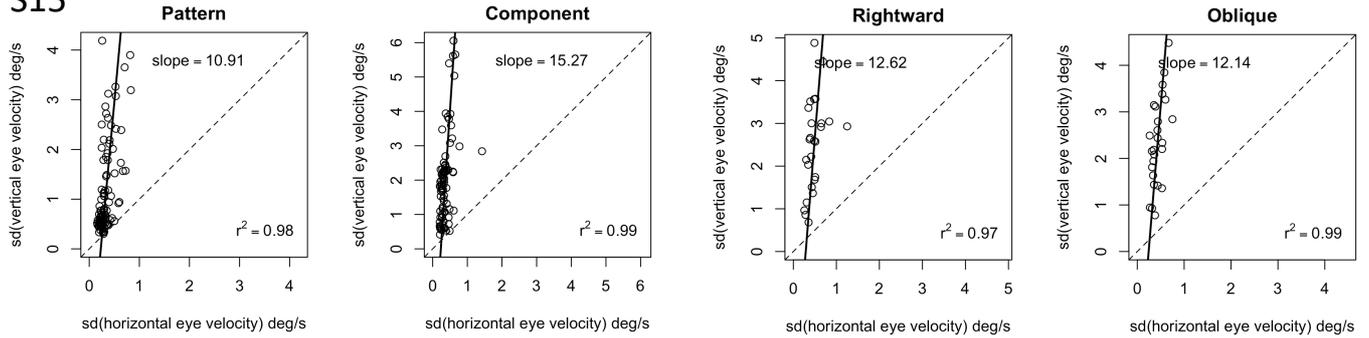
S13



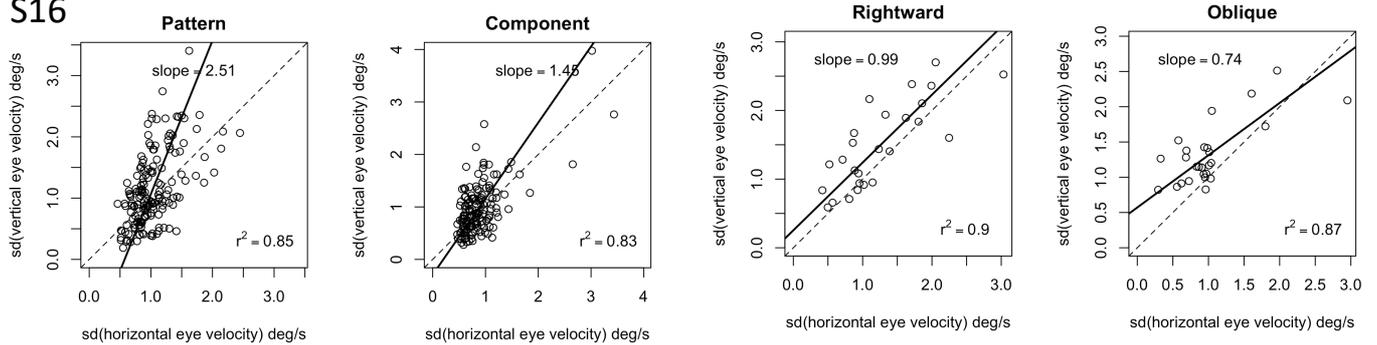
S14



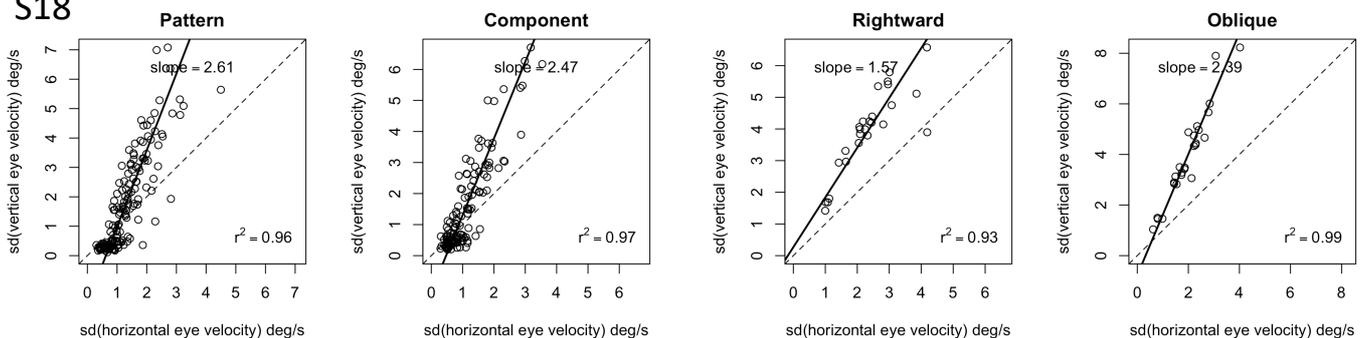
S15



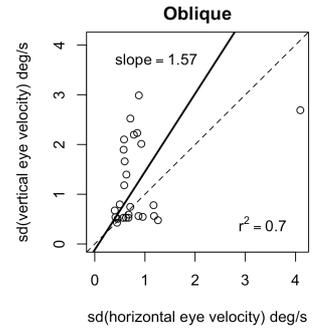
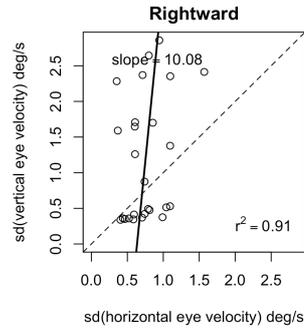
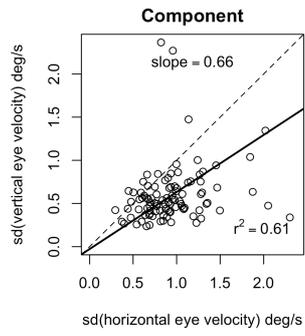
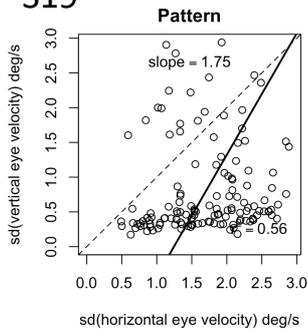
S16



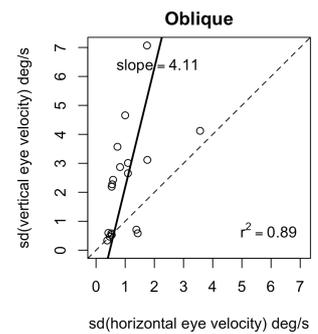
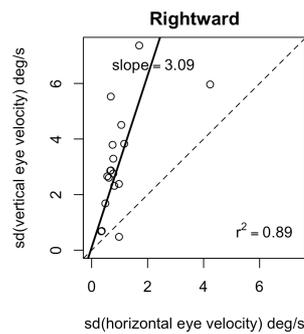
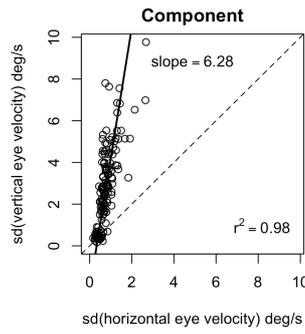
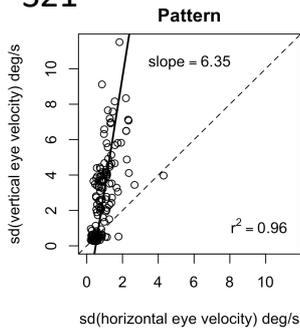
S18



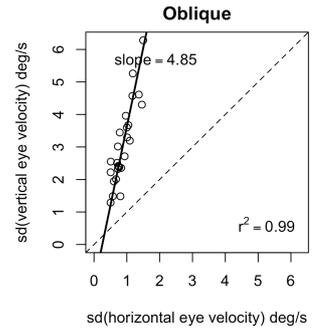
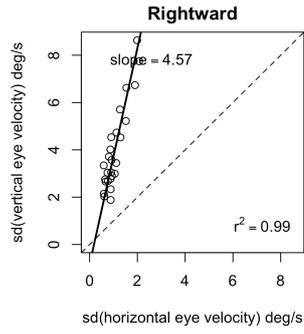
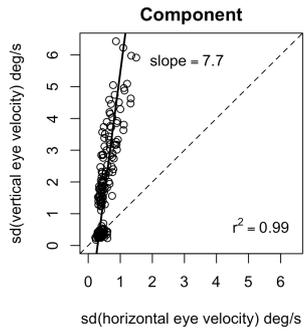
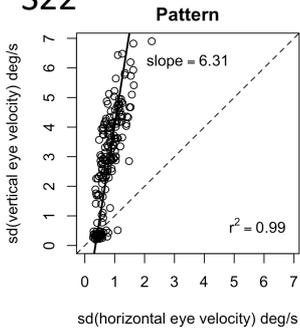
S19



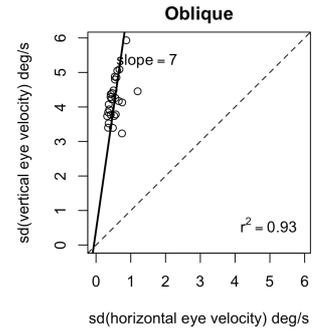
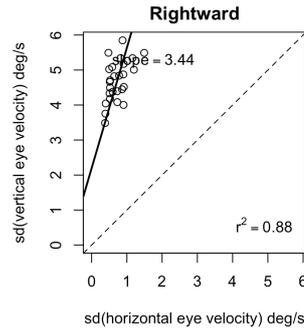
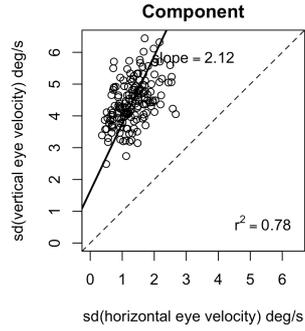
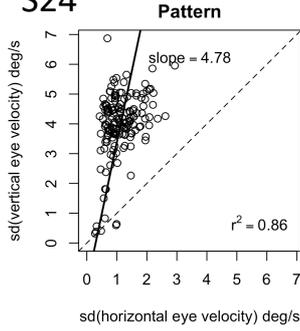
S21



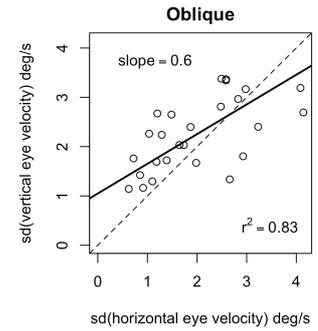
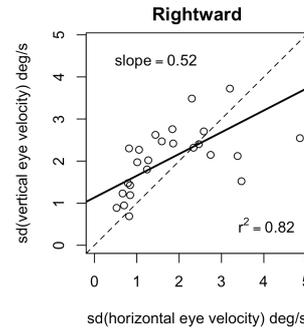
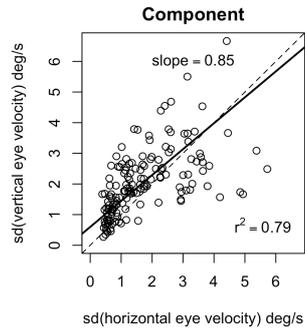
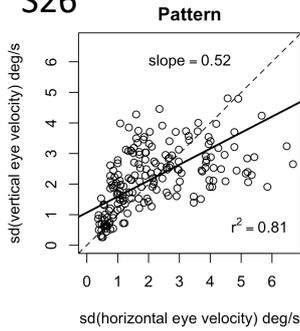
S22



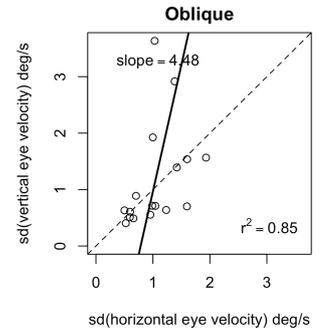
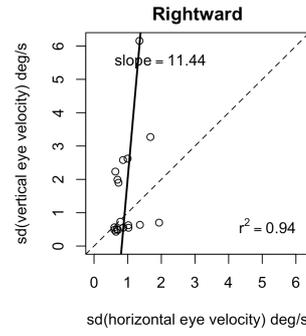
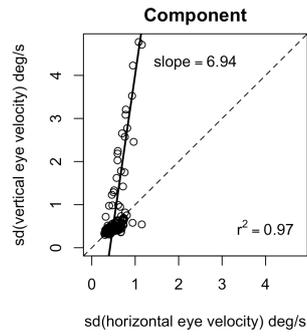
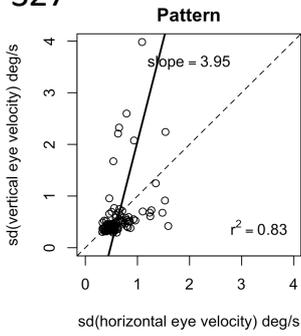
S24



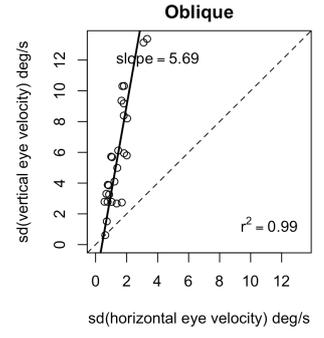
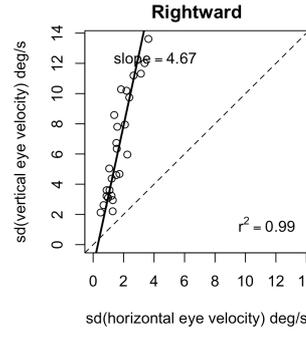
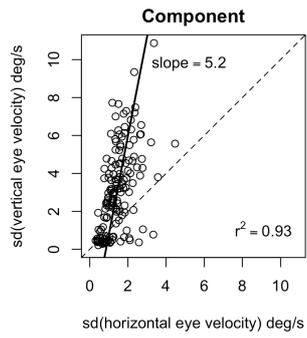
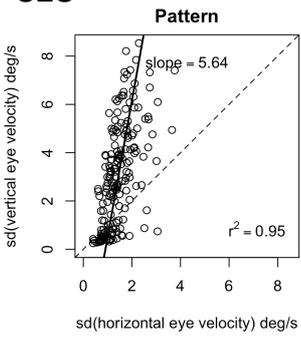
S26



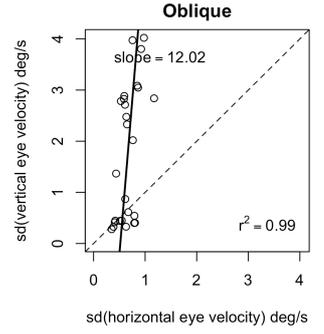
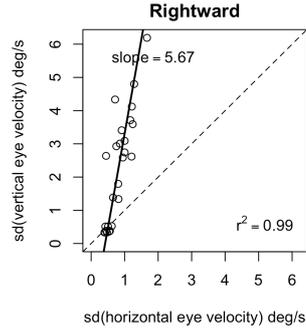
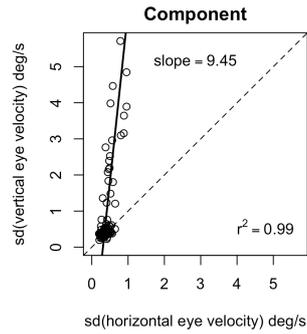
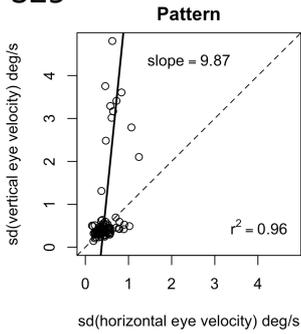
S27



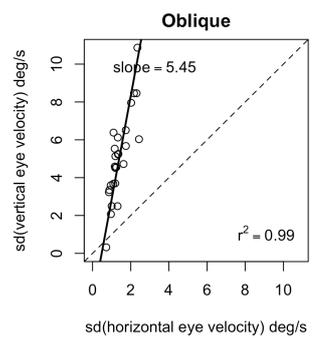
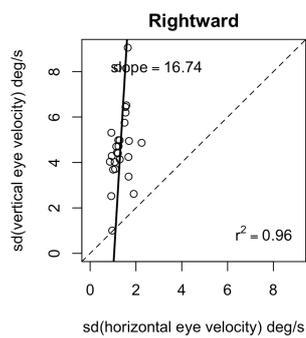
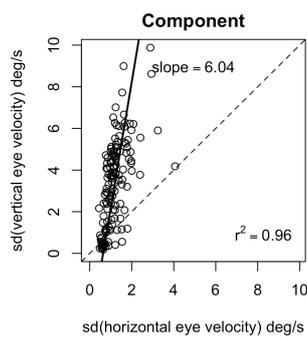
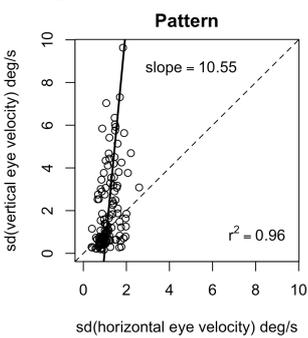
S28



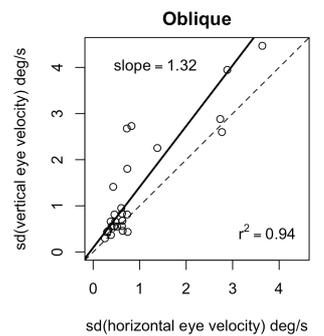
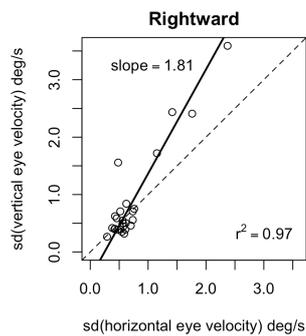
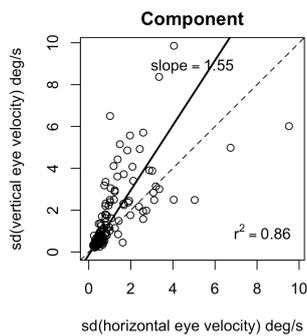
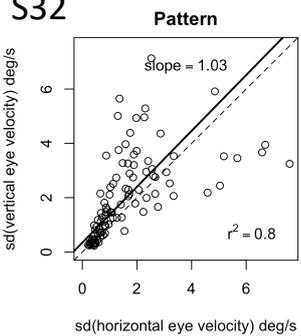
S29



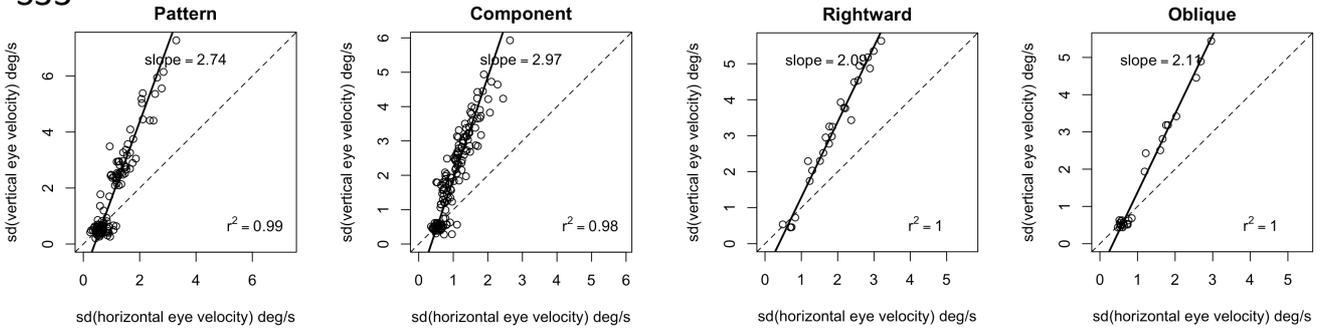
S31



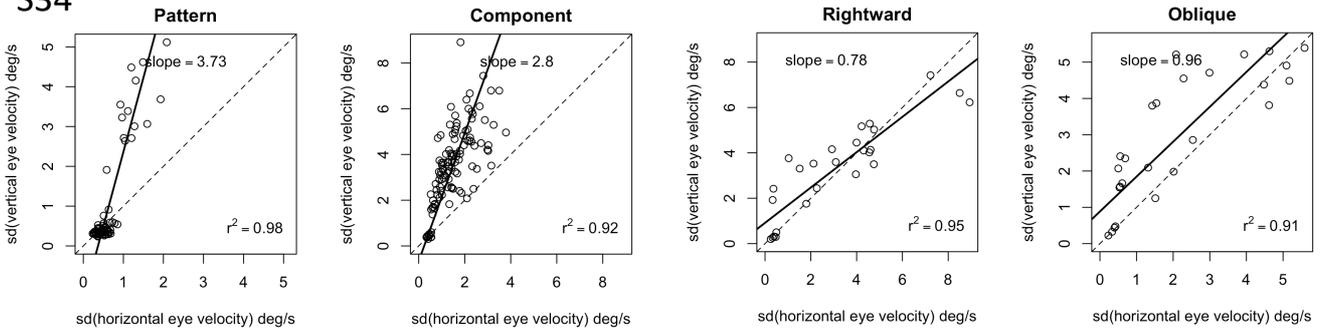
S32



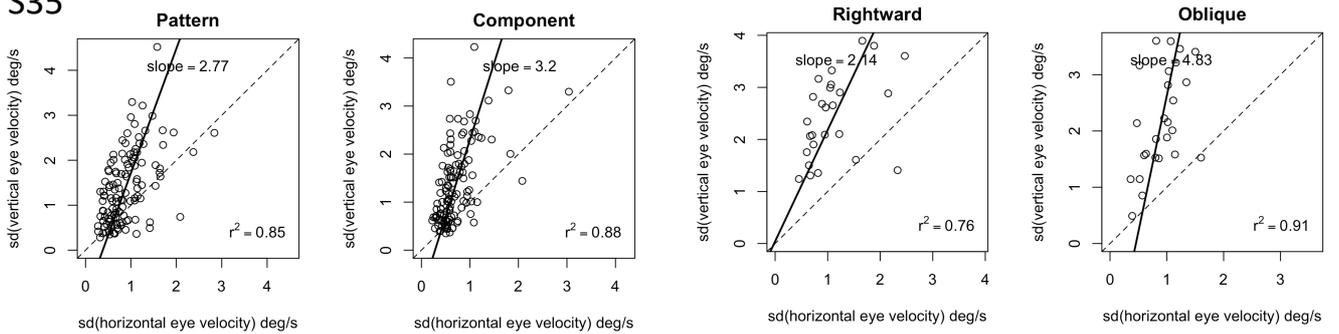
S33



S34



S35



S36

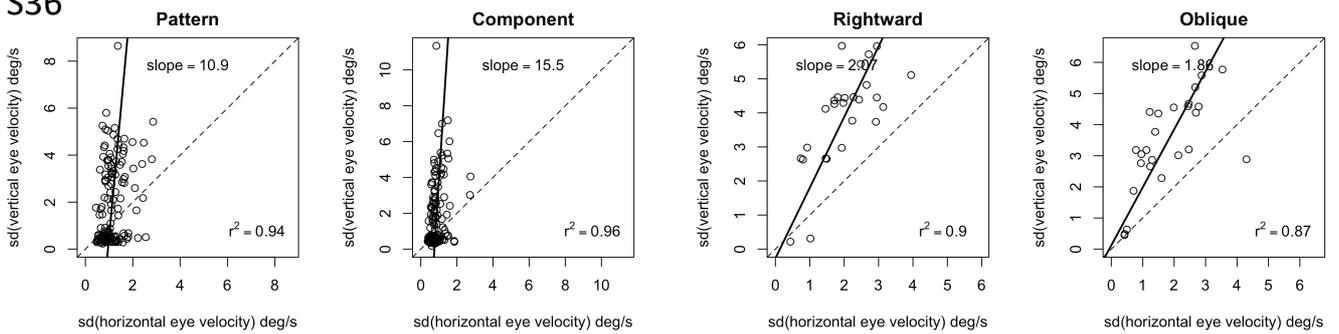


Figure 32: Individual plots of vertical/horizontal eye velocity regression model
 We modeled Deming regression independently for the bistable perception and localizer task independently for each observer. The estimated slope ($\hat{\beta}_1$) and intercept ($\hat{\beta}_0$) are plotted against the identity line (where $\hat{\beta}_1=1$ and $\hat{\beta}_0=0$, dashed line) to evaluate potential bias. A value of $\hat{\beta}_1 > 0$ was indicative of a bias towards vertical eye velocity variations; and reversely, $\hat{\beta}_1 < 0$ of horizontal bias. Model's goodness of fit was also estimated as r^2 (where $0 < r^2 < 1$)

Statistically, we performed two paired t-tests to compare the slope coefficients between Pattern and Components (bistable perception) on the one hand, and Rightward and Oblique (localizer) on the other hand.

4. Results

Eye movement during Pattern and Component percepts

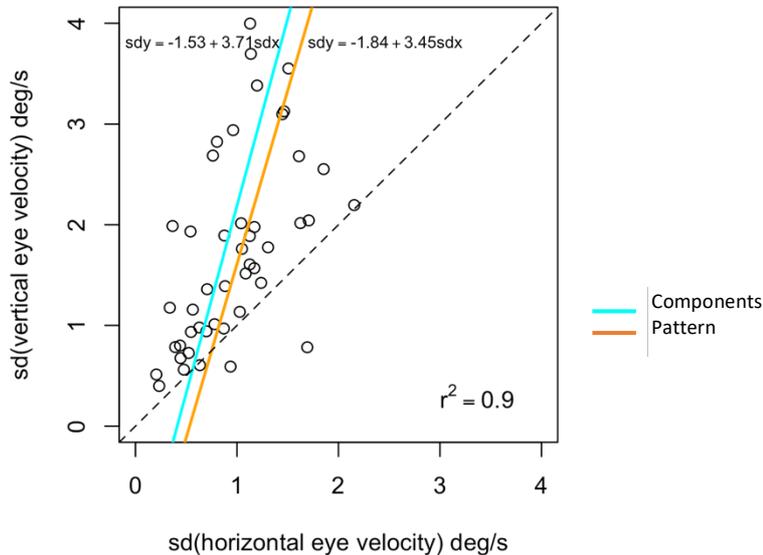


Figure 33: Group-level plot of vertical/horizontal eye velocity regression model during bistable perception
We modeled Deming regression during bistable perception for the group (N=24).
The estimated slope ($\hat{\beta}_1$) and intercept ($\hat{\beta}_0$) are plotted against the identity line (where $\hat{\beta}_1=1$ and $\hat{\beta}_0=0$, dashed line) to evaluate potential bias.
A value of $\hat{\beta}_1 > 0$ was indicative of a bias towards vertical eye velocity variations; and reversely, $\hat{\beta}_1 < 0$ of horizontal bias.
Model's goodness of fit was also estimated as r^2 (where $0 < r^2 < 1$)

Bistable perception: Paired t-test

data: $\hat{\beta}_1$ by percept (intercept: P)

t = 0.68, df = 23, p-value = 0.50 (not significant)

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-10.6 21.0

Eye movement during Rightward and Oblique motion

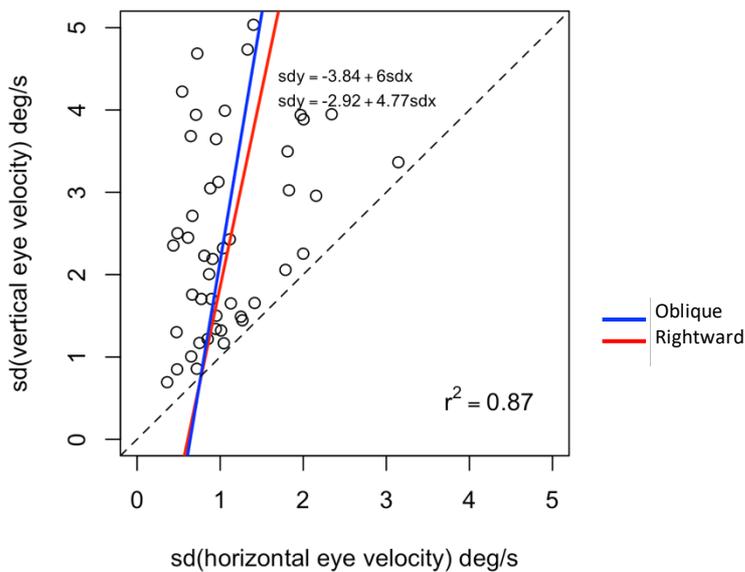


Figure 34: Group-level plot of vertical/horizontal eye velocity regression model during localizer

We modeled Deming regression during the localizer for the group (N=24).

The estimated slope $\hat{\beta}_1$ and intercept $\hat{\beta}_0$ are plotted against the identity line (where $\hat{\beta}_1=1$ and $\hat{\beta}_0=0$, dashed line) to evaluate potential bias.

A value of $\hat{\beta}_1 > 0$ was indicative of a bias towards vertical eye velocity variations; and reversely, $\hat{\beta}_1 < 0$ of horizontal bias.

Model's goodness of fit was also estimated as r^2 (where $0 < r^2 < 1$)

Localizer: Paired t-test

data: $\hat{\beta}_1$ by percept (intercept: P)

t = -0.88, df = 23, p-value = 0.38 (not significant)

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-8.19 3.30

At the group-level, we found no significant bias regarding vertical/horizontal velocity variations, neither across Pattern/Component perceptual states (bistable perception, Fig. 33) nor Rightward/Oblique trials (localizer, Fig. 34). The $\hat{\beta}_1$ coefficient of linear regression $\sigma_y = \hat{\beta}_0 + \hat{\beta}_1\sigma_x + \varepsilon$ did not significantly differ across conditions. Moreover, we confirmed that the residual bias towards making faster eye movements along the vertical axis is generalized across all conditions and motion-related tasks.

The results do not support the hypothesis of a recurrent bias in the eye movements that would affect the neural correlates of perceptual decisions in early visual areas.

5. Conclusion

Both at the subject- and group-levels, statistical analyses did not reveal a significant effect of the involuntary eye movements made just before a perceptual switch, or during bistable perception itself. In a first analysis, we performed individual classification analyses in which we used a reliable estimates of involuntary eye movements (horizontal and vertical variations of eye velocity made before each perceptual switch) to predict the upcoming perceptual state. Statistically, Pattern versus Component decoding did not exceed the acceptability criterion. Secondly, we estimated if the directionality of eye movements reflected the perceived motion during Pattern and Component perceptual states. We did not find such effect, and this was the case in all observers, however we found a general trend in making more eye movements along vertical axis (i.e., towards Component motions) rather than horizontally (i.e., towards Pattern). We believe this might reflect the fact that ocular following response reflects orientation-selective neural mechanisms rather than direction (Sheliga et al, 2015). Therefore, it seems that we can exclude this confounding factor from the potential source of noise in our fMRI data.

This last chapter is a control analysis complementary to the results presented in Chapter 3 on the neural correlates of bistable perception. The current analysis supports that the Pattern/Component contrast estimates that reached significance in direction-selective subdomains of early visual areas (V1, V2 and hMT+) were not biased by eye movement phenomena.

CONCLUSION & PERSPECTIVES

What motion integration reveals about functional hierarchy?

With the objective of studying motion FB to area V1, we used a paradigm of bistable perception that consisted of presenting a fixed visual input that triggers two alternating perceptual states (i.e., a bistable stimulus). The stimulus was dynamic in the sense that it consisted of a correlated spatio-temporal variation (i.e., a moving pattern), but its parameters were fixed (e.g., orientation and contrast of components, temporal and spatial frequencies, etc.), so that any perceptual change in its appearance can be attributed to internal mechanisms rather than variation in the stimulus parameters. We designed an ambiguous moving plaid based on a psychophysical study that aimed at determining the optimal physical parameters to maximize bistable perception, i.e., produce approximately regular perceptual transitions and stable states that were of sufficient duration for fMRI recording. In theory, any perceptual switch is associated with an update of an internal representation.

In order to investigate the behavioral and neural correlates of such a phenomenon, we presented 34 observers with a moving bistable plaid during prolonged presentations. They were asked to fix their gaze on a central target cross to avoid voluntarily switching perception through active sensing. The main task was to dynamically report any perceptual switch by button press. Meanwhile, we measured two behavioral markers (button presses and eye movements) in addition to the BOLD signal variation that we investigated at three different levels of the visual cortical hierarchy: V1, V2 and hMT+. Within each cortical area, we determined direction-selective subdomains based on independent sequences of non-ambiguous moving gratings with directions matching the perceived ones for the bistable plaid. These subdomains were used as regions of interest in the main analysis that consisted of estimating the contrast between Pattern and Component perceptual states. The results demonstrated that differential Pattern vs Component activity can be measured in direction-selective units as early as area V1. Overall, we found that oblique-selective subdomains showed a preference for the component percept, which was perceived as oblique motion, in areas V1, V2 and hMT+. In contrast, rightward-selective subdomains showed a significant preference for the pattern percept, perceived as rightward motion, only in V1. We speculate this might reflect a greater contribution of component-selective units in triggering perceptual alternation. Nevertheless, the relative difference in subdomains' activity during the two different perceptual states was significant in all three visual areas studied.

Subsequently, we analyzed eye movement characteristics associated with bistable perception in order to assess whether systematic effects of eye movement direction could predict the perceptual states. It turned out that directional eye movement velocity was not predictive of the upcoming perceptual switch and nor did it reflect current perceived motion direction. Surprisingly, in most observers, we found more variation of eye velocity vertically (rather than horizontally), and this no matter what motion direction was perceived (Component perceived as oblique; Pattern perceived as rightward). This suggest that the ocular pursuit response reflected stimulus properties (i.e., first-order or Fourier motion signal) rather than perceptual states. This is in accordance with a study by Sheliga et al. (2015) reporting oculomotor responses to bistable plaid that reflected the inhibitory flanks (or end-stopping property) of orientation-selective cells.

FB models of connectivity

Based on the above findings, we hypothesized that higher-level predictions of motion signals are fed back to the earliest cortical relay of vision, area V1. The results motivate further investigation to determine the origin of such a signal. While hMT+ is a potential candidate, an fMRI study examining bistable motion using a structure-from-motion stimulus paradigm, reported evidence of a role of inferior frontal cortex projections to hMT+ in resolving perceptual conflicts between sensory input and top-down predictions (Weilnhammer et al., 2021). Investigating FB activity is challenging, for several reasons. FF processing finds its origin in the outside world so that the experimenter can easily change the configuration of the stimulus and examine how it influences the response. This contrasts with FB signals which are generated by higher levels of the cortex, so that the experimenter has to manipulate the expectation of the subject. This is particularly challenging in non-human subjects and understanding FB at the cellular level will remain so until causative control methods have been developed. Investigation of functional pathways is often limited to measuring activity in the target area, leaving the origin of the signal undetermined. This issue is more specific to FB connections as FB projections cross more hierarchical levels, and are more divergent than FF (Markov et al., 2014b). Although there are ways to compare different models of directed connectivity (e.g., Dynamic Causal Modeling), there is a limited number of areas than can be analyzed simultaneously in such models. Further, functional cortical connectivity is poorly described as a serial one-to-one inter-areal connectivity making the cooperative function of the cortex difficult to pin down, therefore it seems overtly simplistic to only look for one origin of the FB signals. Instead, the cascading top-down signals are presumed to cross multiple cortical relays while inducing local variations of activity in each of them (Fig. 35) (Markov et al., 2014b).

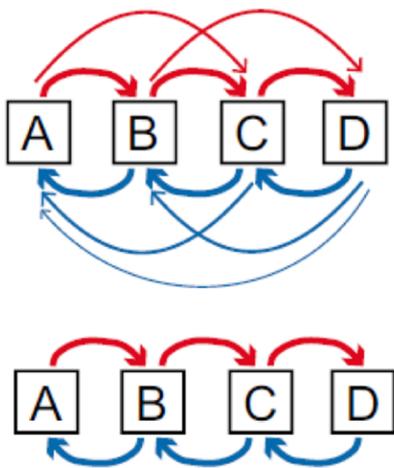


Figure 35: Models of hierarchical connectivity
Lower panel represents the incorrectly assumed serial processing between areas (too simplistic to describe cortical connectivity), where instead each areas project to all upper and lower stream areas (all to all) (upper panel) (Markov et al., 2014b)

Looking for the neural basis of switching phenomena

Functionally, the theory of the canonical cortical microcircuit provides a suitable explanation to account for the selective and local amplification of specific top-down signals over time. For example, it was proposed that the switching mechanisms involved in all-or-none task switching (i.e., during which the attended feature varied across the session) reflect the selective amplification of FB-driven signals in the local cortical circuit accordingly to the new sensory representations at play (Ardid & Wang, 2013). We could imagine a similar scenario leading to perceptual switching in bistable perception. Consequently, the perceptual state would reflect

the dominant amplified top-down prediction signal at a given time, before the reversal of dominance between concurrent top-down signals occur. Testing this hypothesis would require investigating FB functional cortical connectivity with laminar resolution. Indeed, identifying the laminar pattern of FB connectivity involved in motion integration could contribute to further determine the functional specificities of supragranular and infragranular FB streams in hierarchical processing of visual information.

The promise of high-resolution fMRI

The main drawback of investigating bistable perception is that it constrains the experimental conditions to be subjectively defined (based on observer's report) or based on a physiological marker such as predictive eye movements (Wilbertz et al, 2018), rather than relying on experimental conditions that are objectively separated in time (e.g., distinct visual inputs). Therefore, it is of primary importance to define highly selective regions for the feature of interest based on an independent task with unambiguous experimental conditions. Such functional localizers aiming to identify small units with sharp stimulus tuning (e.g., tuned to a particular motion direction in our case) require a fine-grained spatial resolution to improve accuracy. Therefore, our definition of direction-selective subdomains would benefit from being investigated with high-resolution fMRI, which would allow measuring functional connectivity at different cortical depths. Using 7T fMRI, several studies in humans provided evidence that hMT+ contains direction-selective columnar units (Zimmermann et al., 2011; Schneider et al., 2019; Sousa et al., 2021) the activity of which reflects both perceptual state and stimulus physical properties. What remains elusive is how this motion information is fed back to lower hierarchical levels and more specifically with respect to deep vs superficial FB streams.

Functional implications of the Dual Counter Stream

An intriguing feature of anatomical hierarchy is that it is not the laminar location of cells of origin that specifies the FF and FB property of pathways but their *relative proportion* between supragranular and infragranular layers which changes as a smooth function of hierarchical distance. Figure 36 represents SLN (i.e., fraction of supragranular neurons), a general statistical marker of the anatomical cortical hierarchy that led to the Dual Counter stream architecture following retrograde tract-tracing experiments. Importantly, SLN provides a gradient relative to the injected area that measures hierarchical distance (Markov et al, 2014b; Vezoli et al., 2021). Further analyses showed that the change of SLN is due to the supragranular and infragranular FB having very different space constants, so that the short distance supragranular FB decreases with distance faster than the long-distance infragranular FB (Markov et al., 2014b). In addition, an important topographical distinction is made between the point-to-point connectivity reported in the supragranular FB and the diffuse nature of the infragranular FB. Altogether, these anatomical specificities naturally lead to speculations about the functional roles of the two FB pathways (Shipp 2016; Vezoli 2021; Markov et al, 2014b). It was proposed that supragranular FB conveys precision signals (or degree of uncertainty modulating bottom-up and top-down signals) from neighboring areas (e.g., Shipp et al., 2013), which Andy Clark (2013) has likened to a visual processing signal, while infragranular FB would relay top-down prediction signals. Recent laminar high-resolution fMRI in human subjects supports these predictions (see review de Lange et al., 2018).

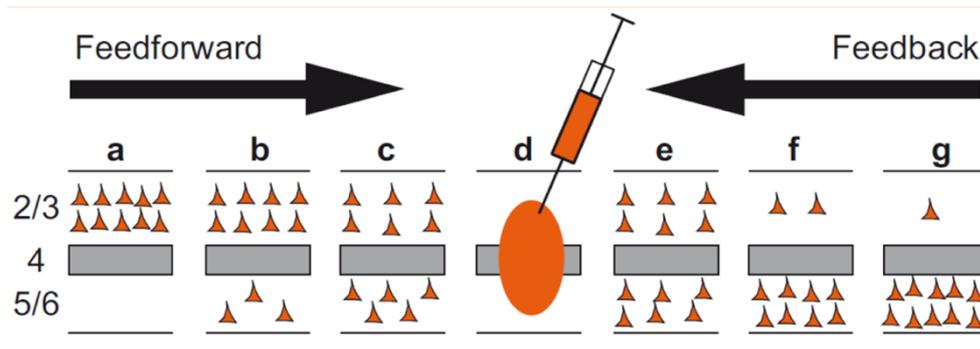


Figure 36: Quantitative retrograde tract-tracing of parent neurons targeting the injected area (d). The laminar distribution of parent neurons in each pathway, referred to as SLN (fraction of supragranular neurons) is determined by high frequency sampling and quantitative analysis of labeling. Supra- and infragranular layer neurons contribute to both FF and FB pathways, and their relative proportion is characteristic for each type of pathway.

For a given injection there is a gradient of SLN of the labeled areas, between purely FF (SLN = 100%, all the parent neurons are in the supragranular layers) to purely FB (SLN = 0%, all the parent neurons in the infragranular layers) and a spectrum of intermediate proportions (Markov et al., 2014b)

Spotting mental representations out of illusory perceptions

Using high-resolution 7T fMRI, Bergmann et al. (2019 bioRxiv) investigated FB profiles of activity across V1 cortical depth during tasks of mental imagery (i.e., generating mental representation given statistical regularities and in the absence of visual input) and illusory perception (i.e., “shaping perception”). They reported a specialization of supragranular FB activity in “real” (or physical) illusory content, whereas the infragranular FB activity was more specific to mental imagery. This study showed that in contrast to mental imagery, illusory content shares information with stimulus-induced visual perception. To further this perspective, a recent review by Maric & Domijan (2022) explored in the literature the notion of “cognitive penetrability”, that represents the influence of top-down predictions at different scales of perceptual processing. They propose that infragranular and supragranular top-down streams do not exert the same influence on the shaping of visual perception. They argue that visual processing is cognitively penetrable particularly to the infragranular FB stream while this is not the case in supragranular layers.

An illusory perception that involves a comparable “shaping perception” type of FB activity lies in the bistable perception of moving plaids. With respect to the Bergmann et al study this could mean that the FB motion signal to area V1 during the bistable motion perception targets the supragranular layers. Investigating perceptual decision from the perspective of low-level visual areas has been linked to the concept of explicit early perception that posits that top-down perceptual inference allows low-level areas to access the “gist of the scene”. Interestingly, a recent study by Papale and collaborators (Papale et al., 2022 bioRxiv) showed evidence of an extremely rapid contextual FB signal to area V1 that shared similar representation with visually driven FF signal, again suggesting that the point-to-point sugragranular FB is involved. Future high-resolution investigation of illusory motion FB to area V1 will be valuable for investigating the precision coding aspects of predictive processes.

The contribution of electrophysiology

Recent fMRI studies with a broad range of paradigms using perceptual inference in humans, have provided insight on multiple visual functions implicating FB (without reaching laminar resolution). More nuanced alternatives to predictive coding have arisen, suggesting control mechanisms enabling to switch between FF and FB perceptual modes of operation over space and time (Heeger et al., 2017). This is supported by invasive laminar recordings in macaque cortex, showing that FF and FB activities seem to occur over distinct temporal windows and at distinct cortical depths (Self et al, 2013; 2019; van Kerkoerle et al., 2017). If we consider area V1 for example, Self and collaborators performed an experiment of boundary detection/surface filling-in showing an early FF signal for boundary detection (+~50ms) in layer 4 and superficial layers, followed by FB activity for surface filling-in (+~100ms) in superficial and deep layers (avoiding layer 4). A comparative temporal distinction was made using a figure/ground segregation experiment in macaque, showing at first, an increased FB signal in response to the figure enhancement, and then a decrease of activity characterizing background suppression (Poort et al., 2016). These findings suggest that successive profiles of neural activity across cortical layers can be functionally linked to different stages of perceptual processing. Illusory perception together with laminar resolved techniques offers numerous possibilities for investigating the function of the distinct FB streams.

Although within my thesis, we focused our investigation on foveal to parafoveal representation of bistable motion (covering 1.5° to 13° degrees of visual angle), one could wonder about a possible differential strength of motion integration signals between central and peripheral vision as we shall discuss in the next section.

Motion representations across visual field

Recalling that all cortical projections to area V1 are feedback, a recent study of the host team mapped out the full complement of FB pathways to retinotopic subdivisions of area V1 (Wang et al., 2022 bioRxiv). The study revealed that peripheral and upper field projections to V1 are stronger from the dorsal stream (where MT/V5 is located), whereas central and lower field projections to V1 are stronger from the ventral stream. These findings alone suggest that MT-to-V1 FB projections might be dominated by peripheral and upper field connections. However, MT/V5-to-V1 FB pathway was the exception to the general rule, by revealing relatively constant strength of connections across eccentricity. Unpublished findings of the team revealed that V1 and V2 peripheral lower field representations receives projections from dorsal stream areas (including MT/V5) over much larger hierarchical distances. With respect to the Dual Counter Stream architecture (Fig. 36), this suggests that the supragranular FB might be weaker in the periphery than in the central representation, a possibility that is being currently investigated by the team. These observations suggest that the proportion of supragranular and infragranular FB could vary with eccentricity in the visual field. This could lead to different sorts of motion FB signals projecting to the peripheral representation of the early visual cortices. Such investigations might corroborate the claim that peripheral vision is mainly dedicated to implicit perception (i.e., processed at the earliest levels of the visual cortical hierarchy) while central vision is linked to “vision with scrutiny”, due to the high acuity conferred by the increased density of cone photoreceptors at the fovea, a form of perceptual inference that would originate from higher-order areas (Tuten & Harmening, 2021; Hochstein & Ahissar, 2002). According to this view, we would expect MT/V5-to-V1 supragranular FB

to be stronger for the central representation than the peripheral one, and specifically involved in contextual perceptual inference of motion information.

Quoting William S. Tuten and Wolf M. Harmening (2021), “most eyes do not have a fovea” per se. Indeed, if we consider the anatomical structure of the fovea as described in some reptiles, fishes, birds, and primates, it occurs that only predators have one, and its structure and functional properties greatly differ across species. This evolution of the retina and foveal representation correlates with the expansion of the supragranular layers in non-avian reptiles and most primates. Altogether, these arguments highlight the functional differences between supragranular and infragranular FB between central and peripheral representation of visual field.

Another intriguing comparison lies between the lower visual field, which contributes more to spatio-temporal processing (thereby expected to be particularly efficient at processing motion signals), while upper visual field is more specific of visual search and attentional shifting (Thomas & Elias, 2011; Kraft et al., 2011). These observations are in line with Previc’s ecological theory of vision (Previc, 1990). This motivates further investigations of the differential strength of FB pathways between lower and upper field representation in early visual areas during a task of bistable motion perception.

Within the framework of FB function and in the particular model of motion integration, our findings suggest that complex representations of motion signals recorded in the earliest cortical stages of visual processing reflect top-down prediction signals. We propose that these signals are fed back from higher-order cortical areas (possibly hMT+) and, being detected as early as area V1, are responsible for the experience of bistable perception, a specific type of perceptual inference that consists of internally generated changes of perceptual state. One plausible explanation of such a phenomenon is that, due to the ambiguity inherent in the visual input, a selective amplification of the most likely predictions arising from areas that compute 2D motion integration is performed locally, within the canonical cortical microcircuit. Investigating such motion integration FB in humans would require a more fine-grained resolution to estimate the profile of activity across cortical depth in area V1, a goal that can be achieved with the use of laminar-resolved fMRI. More generally, this would contribute to reframe the two-stage model of motion integration while integrating an active role of FB signals.

BIBLIOGRAPHY

Adams DL, Sincich LC, Horton JC. Complete pattern of ocular dominance columns in human primary visual cortex. *J Neurosci*. 2007 Sep 26;27(39):10391-403. doi:10.1523/JNEUROSCI.2923-07.2007.

Adelson EH, Movshon JA. Phenomenal coherence of moving visual patterns. *Nature*. 1982 Dec 9;300(5892):523-5. doi:10.1038/300523a0.

Adelson EH, Bergen JR. Spatio-temporal energy models for the perception of motion. *Journal of the Optical Society of America. A, Optics and Image Science*. 1985 Feb;2(2):284-299. doi:10.1364/josaa.2.000284.

Ahmed B, Cordery PM, McLelland D, Bair W, Krug K. Long-range clustered connections within extrastriate visual area V5/MT of the rhesus macaque. *Cereb Cortex*. 2012 Jan;22(1):60-73. doi:10.1093/cercor/bhr072.

Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP. Shared representations for working memory and mental imagery in early visual cortex. *Curr Biol*. 2013 Aug 5;23(15):1427-31. doi:10.1016/j.cub.2013.05.065.

Albright TD. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol*. 1984 Dec;52(6):1106-30. doi:10.1152/jn.1984.52.6.1106.

Angelucci A, Bullier J. Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *J Physiol Paris*. 2003 Mar-May;97(2-3):141-54. doi:10.1016/j.jphysparis.2003.09.001.

Ardid S, Wang XJ. A tweaking principle for executive control: neuronal circuit mechanism for rule-based task switching and conflict resolution. *J Neurosci*. 2013 Dec 11;33(50):19504-17. doi:10.1523/JNEUROSCI.1356-13.2013.

Avants B, Cook PA, McMillan C, Grossman M, Tustison NJ, Zheng Y, Gee JC. Sparse unbiased analysis of anatomical variance in longitudinal imaging. *Med Image Comput Comput Assist Interv*. 2010;13(Pt 1):324-31. doi:10.1007/978-3-642-15705-9_40.

Bahill AT, McDonald JD. Smooth pursuit eye movements in response to predictable target motions. *Vision Res*. 1983;23(12):1573-83. doi:10.1016/0042-6989(83)90171-2.

Baker CL Jr. Central neural mechanisms for detecting second-order motion. *Curr Opin Neurobiol*. 1999 Aug;9(4):461-6. doi:10.1016/S0959-4388(99)80069-5.

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

Barlow HB. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*. 1972;1(4):371-94. doi:10.1068/p010371.

Barone P, Batardiere A, Knoblauch K, Kennedy H. Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J Neurosci*. 2000 May 1;20(9):3263-81. doi:10.1523/JNEUROSCI.20-09-03263.2000.

- Barzegaran E, Plomp G. Four concurrent feedforward and feedback networks with different roles in the visual cortical hierarchy. *PLoS Biol.* 2022 Feb 10;20(2):e3001534. doi:10.1371/journal.pbio.3001534.
- Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. Canonical microcircuits for predictive coding. *Neuron.* 2012 Nov 21;76(4):695-711. doi:10.1016/j.neuron.2012.10.038.
- Bastos AM, Vezoli J, Bosman CA, Schoffelen JM, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P. Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron.* 2015 Jan 21;85(2):390-401. doi:10.1016/j.neuron.2014.12.018.
- Bates D, Maechler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- Bayerl P, Neumann H. Disambiguating visual motion through contextual feedback modulation. *Neural Comput.* 2004 Oct;16(10):2041-66. doi:10.1162/0899766041732404.
- Beck J, Prazdny K, Ivry R. The perception of transparency with achromatic colors. *Percept Psychophys.* 1984 May;35(5):407-22. doi:10.3758/bf03203917.
- Behzadi Y, Restom K, Liao J, Liu TT. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage.* 2007 Aug 1;37(1):90-101. doi:10.1016/j.neuroimage.2007.04.042.
- Bell AJ, Sejnowski TJ. The "independent components" of natural scenes are edge filters. *Vision Res.* 1997 Dec;37(23):3327-38. doi:10.1016/s0042-6989(97)00121-1.
- Bergmann J, Morgan AT, Muckli L. Two distinct feedback codes in V1 for 'real' and 'imaginary' internal experiences. *bioRxiv* 664870, 2019; doi:10.1101/664870.
- Berman RA, Wurtz RH. Signals conveyed in the pulvinar pathway from superior colliculus to cortical area MT. *J Neurosci.* 2011 Jan 12;31(2):373-84. doi:10.1523/JNEUROSCI.4738-10.2011.
- Berzhanskaya J, Grossberg S, Mingolla E. Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spat Vis.* 2007;20(4):337-95. doi:10.1163/156856807780919000.
- Bialek W, De Ruyter RR, Steveninck V, Tishby N. Efficient representation as a design principle for neural coding and computation. 2006 IEEE International Symposium on Information Theory. 2006, pp. 659-663, doi:10.1109/ISIT.2006.261867.
- Binzegger T, Douglas RJ, Martin KA. Topology and dynamics of the canonical circuit of cat V1. *Neural Netw.* 2009 Oct;22(8):1071-8. doi:10.1016/j.neunet.2009.07.011.
- Blake R & Fox R. Adaptation to invisible gratings and the site of binocular rivalry suppression. *Nature*, 1974, May 31; 249 (456), 488-490.

Brecht M, Roth A, Sakmann B. Dynamic receptive fields of reconstructed pyramidal cells in layers 3 and 2 of rat somatosensory barrel cortex. *J Physiol.* 2003 Nov 15;553(Pt 1):243-65. doi: 10.1113/jphysiol.2003.044222.

Breese BB. On inhibition. *The Psychological Review: Monograph Supplements*, 1899; 3 (1), i-65, doi:10.1037/h0092990.

Bogadhi AR, Montagnini A, Mamassian P, Perrinet LU, Masson GS. Pursuing motion illusions: a realistic oculomotor framework for Bayesian inference. *Vision Res.* 2011 Apr 22;51(8):867-80. doi:10.1016/j.visres.2010.10.021.

Bosman CA, Schoffelen JM, Brunet N, Oostenveld R, Bastos AM, Womelsdorf T, Rubehn B, Stieglitz T, De Weerd P, Fries P. Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron.* 2012 Sep 6;75(5):875-88. doi:10.1016/j.neuron.2012.06.037.

Bourne JA, Morrone MC. Plasticity of Visual Pathways and Function in the Developing Brain: Is the Pulvinar a Crucial Player? *Front Syst Neurosci.* 2017 Feb 8;11:3. doi:10.3389/fnsys.2017.00003.

Breese BB. Binocular rivalry. *Psychol. Rev.*, 1909, 16, 410-415.

Bremmer F, Kubischik M, Hoffmann KP, Krekelberg B. Neural dynamics of saccadic suppression. *J Neurosci.* 2009 Oct 7;29(40):12374-83. doi:10.1523/JNEUROSCI.2908-09.2009.

Brindle K. New approaches for imaging tumour responses to treatment. *Nat Rev Cancer.* 2008 Feb;8(2):94-107. doi:10.1038/nrc2289.

Brown H, Friston KJ. Free-energy and illusions: the cornsweet effect. *Front Psychol.* 2012 Feb 29;3:43. doi:10.3389/fpsyg.2012.00043.

Brych M, Murali S, Händel B. The Role of Blinks, Microsaccades and their Retinal Consequences in Bistable Motion Perception. *Front Psychol.* 2021 Apr 8;12:647256. doi:10.3389/fpsyg.2021.647256.

Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R. Laminar differences in gamma and alpha coherence in the ventral stream. *Proc Natl Acad Sci U S A.* 2011 Jul 5;108(27):11262-7. doi:10.1073/pnas.1011284108.

Bullier J, 2006. What is Feed Back? In: van Hemmen JL, Sejnowski TJ (Eds.), *23 Problems in Systems Neuroscience*. Oxford University Press USA, pp. 103–132.

Carl JR, Gellman RS. Human smooth pursuit: stimulus-dependent responses. *J Neurophysiol.* 1987 May;57(5):1446-63. doi:10.1152/jn.1987.57.5.1446.

Cavanaugh JR, Bair W, Movshon JA. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J Neurophysiol.* 2002 Nov;88(5):2530-46. doi:10.1152/jn.00692.2001.

Chalk M, Marre O, Tkačik G. Toward a unified theory of efficient, predictive, and sparse coding. *Proc Natl Acad Sci U S A*. 2018 Jan 2;115(1):186-191. doi:10.1073/pnas.1711114115.

Chen KJ, Sheliga BM, Fitzgibbon EJ, Miles FA. Initial ocular following in humans depends critically on the fourier components of the motion stimulus. *Ann N Y Acad Sci*. 2005 Apr;1039:260-71. doi:10.1196/annals.1325.025.

Cheng K, Waggoner RA, Tanaka K. Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron*. 2001 Oct 25;32(2):359-74. doi:10.1016/s0896-6273(01)00477-9.

Chubb C, Sperling G. Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J Opt Soc Am A*. 1988 Nov;5(11):1986-2007. doi:10.1364/josaa.5.001986.

Chubb C, McGowan J, Sperling G, Werkhoven P. Non-Fourier motion analysis. *Ciba Found Symp*. 1994;184:193-205; discussion 206-10, 269-71. doi:10.1002/9780470514610.ch10.

Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci*. 2013 Jun;36(3):181-204. doi:10.1017/S0140525X12000477.

Colman AM, *A Dictionary of Psychology*, "Schröder Staircase", Oxford 2008, online 4th edition 2014. doi:10.1093/acref/9780199534067.001.0001.

Cossell L, Iacaruso MF, Muir DR, Houlton R, Sader EN, Ko H, Hofer SB, Mrsic-Flogel TD. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*. 2015 Feb 19;518(7539):399-403. doi: 10.1038/nature14182.

Cottareau B, Lorenceau J, Gramfort A, Clerc M, Thirion B, Baillet S. Phase delays within visual cortex shape the response to steady-state visual stimulation. *Neuroimage*. 2011 Feb 1;54(3):1919-29. doi:10.1016/j.neuroimage.2010.10.004.

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996 Jun;29(3):162-73. doi:10.1006/cbmr.1996.0014.

de Brouwer S, Yuksel D, Blohm G, Missal M, Lefèvre P. What triggers catch-up saccades during visual tracking? *J Neurophysiol*. 2002 Mar;87(3):1646-50. doi:10.1152/jn.00432.2001.

de Lange FP, Heilbron M, Kok P. How Do Expectations Shape Perception? *Trends Cogn Sci*. 2018 Sep;22(9):764-779. doi:10.1016/j.tics.2018.06.002.

De Valois RL, Yund EW, Hepler N. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res*. 1982a;22(5):531-44. doi:10.1016/0042-6989(82)90112-2.

De Valois RL, Albrecht DG, Thorell LG. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res*. 1982b;22(5):545-59. doi:10.1016/0042-6989(82)90113-4.

Devinck F, & Knoblauch K. Central mechanisms of perceptual filling-in. *Current Opinion in Behavioral Sciences*, 30 (C). doi:10.1016/j.cobeha.2019.08.003.

Dimova KD, Denham MJ. A model of plaid motion perception based on recursive Bayesian integration of the 1-D and 2-D motions of plaid features. *Vision Res.* 2010 Mar 17;50(6):585-97. doi:10.1016/j.visres.2010.01.004.

Diogo AC, Soares JG, Koulakov A, Albright TD, Gattass R. Electrophysiological imaging of functional architecture in the cortical middle temporal visual area of *Cebus apella* monkey. *J Neurosci.* 2003 May 1;23(9):3881-98. doi:10.1523/JNEUROSCI.23-09-03881.2003.

Douglas RJ, Martin KA, Whitteridge D. A Canonical Microcircuit for Neocortex. *Neural Comput* 1989; 1 (4): 480–488. doi:10.1162/neco.1989.1.4.480.

Douglas RJ, Martin KA. A functional microcircuit for cat visual cortex. *J Physiol.* 1991;440:735-69. doi:10.1113/jphysiol.1991.sp018733.

Douglas RJ, Koch C, Mahowald M, Martin KA, Suarez HH. Recurrent excitation in neocortical circuits. *Science.* 1995 Aug 18;269(5226):981-5. doi:10.1126/science.7638624.

Duchemin A, Privat M, Sumbre G. Fourier Motion Processing in the Optic Tectum and Pretectum of the Zebrafish Larva. *Front Neural Circuits.* 2022 Jan 7;15:814128. doi:10.3389/fncir.2021.814128.

Dürsteler MR, Wurtz RH. Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST. *J Neurophysiol.* 1988 Sep;60(3):940-65. doi:10.1152/jn.1988.60.3.940.

Einhäuser W, Stout J, Koch C, Carter O. Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proc Natl Acad Sci U S A.* 2008 Feb 5;105(5):1704-9. doi:10.1073/pnas.0707727105.

Emerson RC, Bergen JR, Adelson EH. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Res.* 1992 Feb;32(2):203-18. doi:10.1016/0042-6989(92)90130-b.

Enoksson P. binocular rivalry and monocular dominance studied with optokinetic nystagmus. *Acta Ophthalmol (Copenh).* 1963;41:544-63. doi:10.1111/j.1755-3768.1963.tb03568.x.

Farid H, Simoncelli EP, Bravo MJ, Schrater PR. Effect of contrast and period on perceived coherence of moving square-wave plaids. *Investigative Ophthalmology and Visual Science Supplement (ARVO)*, 1995, 36:S-51.

Feldman H, Friston KJ. Attention, uncertainty, and free-energy. *Front Hum Neurosci.* 2010 Dec 2;4:215. doi:10.3389/fnhum.2010.00215.

Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex.* 1991 Jan-Feb;1(1):1-47. doi:10.1093/cercor/1.1.1-a.

- Fennema CL, Thompson WB. Velocity determination in scenes containing several moving images. *Comp Graph Image Proc.* 1979;9:301–315.
- Ferrera VP, Wilson HR. Perceived direction of moving two-dimensional patterns. *Vision Res.* 1990;30:273–287.
- Ferrera VP, Wilson HR. Perceived speed of moving two-dimensional patterns. *Vision Res.* 1991;31(5):877-93. doi:10.1016/0042-6989(91)90154-w.
- Ferrera VP, Lisberger SG. Attention and target selection for smooth pursuit eye movements. *J Neurosci.* 1995 Nov;15(11):7472-84. doi:10.1523/JNEUROSCI.15-11-07472.1995.
- Ferster D, Chung S, Wheat H. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature.* 1996 Mar 21;380(6571):249-52. doi:10.1038/380249a0.
- Fisken RA, Garey LJ, Powell TP. The intrinsic, association and commissural connections of area 17 on the visual cortex. *Philos Trans R Soc Lond B Biol Sci.* 1975 Nov 20;272(919):487-536. doi:10.1098/rstb.1975.0099.
- Fonov VS, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL and BDCG, Unbiased average age-appropriate atlases for pediatric studies, *NeuroImage*, Volume 54, Issue 1, January 2011, ISSN 1053–8119, doi:10.1016/j.neuroimage.2010.07.033.
- Fonov VS, Evans AC, McKinstry RC, Almli CR and Collins DL, Unbiased nonlinear average age-appropriate brain templates from birth to adulthood, *NeuroImage*, Volume 47, Supplement 1, July 2009, Page S102 Organization for Human Brain Mapping 2009 Annual Meeting. doi:10.1016/S1053-8119(09)70884-5.
- Friston K. Hierarchical models in the brain. *PLoS Comput Biol.* 2008 Nov;4(11):e1000211. doi:10.1371/journal.pcbi.1000211.
- Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci.* 2010 Feb;11(2):127-38. doi:10.1038/nrn2787.
- Friston K. The history of the future of the Bayesian brain. *Neuroimage.* 2012 Aug 15;62(2):1230-3. doi:10.1016/j.neuroimage.2011.10.004.
- Friston K. Does predictive coding have a future? *Nat Neurosci.* 2018 Aug;21(8):1019-1021. doi:10.1038/s41593-018-0200-7. Erratum in: *Nat Neurosci.* 2019 Jan;22(1):144.
- Friston K. Waves of prediction. *PLoS Biol.* 2019 Oct 3;17(10):e3000426. doi:10.1371/journal.pbio.3000426.
- Frost A, Niemeier M. Suppression and reversal of motion perception around the time of the saccade. *Front Syst Neurosci.* 2015 Oct 31;9:143. doi:10.3389/fnsys.2015.00143.
- Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature.* 1992 Nov 26;360(6402):343-6. doi:10.1038/360343a0.

Georgeson MA, Harris MG. The temporal range of motion sensing and motion perception. *Vision Res.* 1990;30(4):615-9. doi:10.1016/0042-6989(90)90072-s.

Gerardin P, Abbatecola C, Devinck F, Kennedy H, Dojat M, Knoblauch K. Neural circuits for long-range color filling-in. *Neuroimage.* 2018 Nov 1;181:30-43. doi:10.1016/j.neuroimage.2018.06.083.

Gilbert CD, Li W. Top-down influences on visual processing. *Nat Rev Neurosci.* 2013 May;14(5):350-63. doi:10.1038/nrn3476.

Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage.* 2009 Oct 15;48(1):63-72. doi:10.1016/j.neuroimage.2009.06.060.

Greve DN, Brown GG, Mueller BA, Glover G, Liu TT; Function Biomedical Research Network. A survey of the sources of noise in fMRI. *Psychometrika.* 2013 Jul;78(3):396-416. doi:10.1007/s11336-012-9294-0. Erratum in: *Psychometrika.* 2013 Jul;78(3):395.

Guan SC, Zhang SH, Zhang YC, Tang SM, Yu C. Plaid Detectors in Macaque V1 Revealed by Two-Photon Calcium Imaging. *Curr Biol.* 2020 Mar 9;30(5):934-940.e3. doi:10.1016/j.cub.2020.01.005.

Guan SC, Ju NS, Tao L, Tang SM, Yu C. Functional organization of spatial frequency tuning in macaque V1 revealed with two-photon calcium imaging. *Prog Neurobiol.* 2021 Oct;205:102120. doi:10.1016/j.pneurobio.2021.102120.

Guo K, Benson PJ, Blakemore C. Pattern motion is present in V1 of awake but not anaesthetized monkeys. *Eur J Neurosci.* 2004 Feb;19(4):1055-66. doi:10.1111/j.1460-9568.2004.03212.x.

Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. *Nature.* 2009 Apr 2;458(7238):632-5. doi:10.1038/nature07832.

Heeger DJ. Theory of cortical function. *Proc Natl Acad Sci U S A.* 2017 Feb 21;114(8):1773-1782. doi:10.1073/pnas.1619788114.

Henry GH, Mustari MJ, Bullier J. Different geniculate inputs to B and C cells of cat striate cortex. *Exp Brain Res.* 1983;52(2):179-89. doi:10.1007/BF00236626.

Hilgetag CC, O'Neill MA, Young MP. On hierarchies: *Science.* 1996 Feb 9;271(5250):777b. doi:10.1126/science.271.5250.777b.

Hinton GE. Learning multiple layers of representation. *Trends Cogn Sci.* 2007 Oct;11(10):428-34. doi:10.1016/j.tics.2007.09.004

Hochstein S, Ahissar M. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron.* 2002 Dec 5;36(5):791-804. doi:10.1016/s0896-6273(02)01091-7.

Hohwy J. *The Predictive Mind.* Oxford, 2013; online edn, Oxford Academic, 23 Jan 2014. doi:10.1093/acprof:oso/9780199682737.001.0001.

Hohwy J. Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Conscious Cogn.* 2017 Jan;47:75-85.

Hosmer DW Jr, Lemeshow S. *Applied Logistic Regression*. John Wiley & Sons, 2004. p. 177.

Huang T, Chen X, Jiang J, Zhen Z, Liu J. A probabilistic atlas of the human motion complex built from large-scale functional localizer data. *Hum Brain Mapp.* 2019 Aug 15;40(12):3475-3487. doi:10.1002/hbm.24610.

Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol.* 1962 Jan;160(1):106-54. doi:10.1113/jphysiol.1962.sp006837.

Hubel DH, Wiesel TN. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J Neurophysiol.* 1965 Mar;28:229-89. doi:10.1152/jn.1965.28.2.229.

Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. *J Physiol.* 1968 Mar;195(1):215-43. doi:10.1113/jphysiol.1968.sp008455.

Hubel DH, Wiesel TN. Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *J Comp Neurol.* 1974 Dec 1;158(3):295-305. doi:10.1002/cne.901580305.

Hubel DH, Wiesel TN, LeVay S. Functional architecture of area 17 in normal and monocularly deprived macaque monkeys. *Cold Spring Harb Symp Quant Biol.* 1976;40:581-9. doi:10.1101/sqb.1976.040.01.054.

Hubel DH, Wiesel TN. Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B Biol Sci.* 1977 Jul 28;198(1130):1-59. doi:10.1098/rspb.1977.0085.

Hubel DH. *Eye, brain, and vision*. Scientific American Library/Scientific American Books, 1995.

Huguet G, Rinzel J, Hupé JM. Noise and adaptation in multistable perception: noise drives when to switch, adaptation determines percept choice. *J Vis.* 2014 Mar 13;14(3):19. doi:10.1167/14.3.19.

Hupé JM, Rubin N. Dynamics of bi-stable perception of plaids. *Journal of vision*, 2001; 1(3), 157a. doi:10.1167/1.3.157.

Hupé JM, Rubin N. The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Res.* 2003 Mar;43(5):531-48. doi:10.1016/s0042-6989(02)00593-x.

Hupé JM, Lamirel C, Lorenceau J. Pupil dilation does not predict subsequent stability in perceptual rivalry. *Proc Natl Acad Sci U S A.* 2008 Jul 15;105(28):E43; author reply E44. doi:10.1073/pnas.0803456105.

Hupé JM, Lamirel C, Lorenceau J. Pupil dynamics during bistable motion perception. *J Vis.* 2009 Jul 15;9(7):10. doi:10.1167/9.7.10.

Hupé JM, Pressnitzer D. The initial phase of auditory and visual scene analysis. *Philos Trans R Soc Lond B Biol Sci.* 2012 Apr 5;367(1591):942-53. doi:10.1098/rstb.2011.0368.

Hupé JM, Signorelli CM, Alais D. Two paradigms of bistable plaid motion reveal independent mutual inhibition processes. *J Vis.* 2019 Apr 1;19(4):5. doi:10.1167/19.4.5.

Ibbotson MR, Crowder NA, Cloherty SL, Price NS, Mustari MJ. Saccadic modulation of neural responses: possible roles in saccadic suppression, enhancement, and time compression. *J Neurosci.* 2008 Oct 22;28(43):10952-60. doi:10.1523/JNEUROSCI.3950-08.2008.

Ilg UJ. Slow eye movements. *Prog Neurobiol.* 1997 Oct;53(3):293-329. doi:10.1016/s0301-0082(97)00039-7.

Ilg UJ. Commentary: smooth pursuit eye movements: from low-level to high-level vision. *Prog Brain Res.* 2002;140:279-98. doi:10.1016/s0079-6123(02)40057-x.

Ilg UJ, Thier P. Visual tracking neurons in primate area MST are activated by smooth-pursuit eye movements of an "imaginary" target. *J Neurophysiol.* 2003 Sep;90(3):1489-502. doi:10.1152/jn.00272.2003.

Iqbal ST, Zheng XS, Bailey BP. Task-evoked pupillary response to mental workload in human-computer interaction. *Proceedings of the Conference on Human Factors in Computing Systems (CHI'04), Vienna, 2004.* 1477–1480.

James W. *The Principles of Psychology*, in two volumes. New York: Henry Holt and Company, 1890.

Johnson EC, Jones DL, Ratnam R. A minimum-error, energy-constrained neural code is an instantaneous-rate code. *J Comput Neurosci.* 2016 Apr;40(2):193-206. doi:10.1007/s10827-016-0592-x.

Jorstad NL, Close J, Johansen N, Yanny AM, Barkan ER, Travaglini KJ, Bertagnolli D, Campos J, Casper T, Crichton K, Dee N, Ding S-L, Gelfand E, Goldy J, Hirschstein D, Kroll M, Kunst M, Lathia K, Long B, Martin N, McMillen D, Pham T, Rimorin C, Ruiz A, Shapovalova N, Shehata S, Siletti K, Somasundaram S, Sulc J, Tieu M, Torkelson A, Tung H, Ward K, Callaway EM, Hof PR, Keene CD, Levi BP, Linnarsson S, Mitra PP, Smith K, Hodge RD, Bakken TE, Lein ES. Transcriptomic cytoarchitecture reveals principles of human neocortex organization. *bioRxiv* 2022.11.06.515349. doi:10.1101/2022.11.06.515349.

Ju NS, Guan SC, Tao L, Tang SM, Yu C. Orientation Tuning and End-stopping in Macaque V1 Studied with Two-photon Calcium Imaging. *Cereb Cortex.* 2021 Mar 5;31(4):2085-2097. doi:10.1093/cercor/bhaa346.

Kanizsa G. *Organization in Vision: Essays on Gestalt Perception.* New York: Praeger, 1979.

- Keller EL, Heinen SJ. Generation of smooth-pursuit eye movements: neuronal mechanisms and pathways. *Neurosci Res.* 1991 Jul;11(2):79-107. doi:10.1016/0168-0102(91)90048-4.
- Kennedy H, Bullier J. A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. *J Neurosci.* 1985 Oct;5(10):2815-30. doi:10.1523/JNEUROSCI.05-10-02815.1985.
- Kety SS, Schmidt CF. the nitrous oxide method for the quantitative determination of cerebral blood flow in man: theory, procedure and normal values. *J Clin Invest.* 1948 Jul;27(4):476-83. doi:10.1172/JCI101994.
- Khawaja FA, Tsui JM, Pack CC. Pattern motion selectivity of spiking outputs and local field potentials in macaque visual cortex. *J Neurosci.* 2009 Oct 28;29(43):13702-9. doi:10.1523/JNEUROSCI.2844-09.2009.
- Kim J, Wilson HR. Dependence of plaid motion coherence on component grating directions. *Vision Res.* 1993 Dec;33(17):2479-89. doi:10.1016/0042-6989(93)90128-j.
- Klink PC, van Ee R, van Wezel RJ. General validity of Levelt's propositions reveals common computational mechanisms for visual rivalry. *PLoS One.* 2008;3(10):e3473. doi:10.1371/journal.pone.0003473. Erratum in: *PLoS One.* 2016;11(1):e0147454.
- Kok P, Failing MF, de Lange FP. Prior expectations evoke stimulus templates in the primary visual cortex. *J Cogn Neurosci.* 2014 Jul;26(7):1546-54. doi:10.1162/jocn_a_00562.
- Kok P, Bains LJ, van Mourik T, Norris DG, de Lange FP. Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Curr Biol.* 2016 Feb 8;26(3):371-6. doi:10.1016/j.cub.2015.12.038.
- 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 Aug;60(2):580-603. doi:10.1152/jn.1988.60.2.580.
- Komatsu H, Wurtz RH. Modulation of pursuit eye movements by stimulation of cortical areas MT and MST. *J Neurophysiol.* 1989 Jul;62(1):31-47. doi:10.1152/jn.1989.62.1.31.
- Kraft A, Sommer WH, Schmidt S, Brandt SA. Dynamic upper and lower visual field preferences within the human dorsal frontoparietal attention network. *Hum Brain Mapp.* 2011 Jul;32(7):1036-49. doi:10.1002/hbm.21087.
- Krauzlis RJ, Zivotofsky AZ, Miles FA. Target selection for pursuit and saccadic eye movements in humans. *J Cogn Neurosci.* 1999 Nov;11(6):641-9. doi:10.1162/089892999563706.
- Krauzlis RJ. Recasting the smooth pursuit eye movement system. *J Neurophysiol.* 2004 Feb;91(2):591-603. doi:10.1152/jn.00801.2003.
- Krauzlis RJ. The control of voluntary eye movements: new perspectives. *Neuroscientist.* 2005 Apr;11(2):124-37. doi:10.1177/1073858404271196.

Kriegeskorte N, Cusack R, Bandettini P. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatio-temporal filter? *Neuroimage*. 2010 Feb 1;49(3):1965-76. doi:10.1016/j.neuroimage.2009.09.059.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*. 2009 May;12(5):535-40. doi:10.1038/nn.2303.

Laing CR, Chow CC. A spiking neuron model for binocular rivalry. *J Comput Neurosci*. 2002 Jan-Feb;12(1):39-53. doi:10.1023/a:1014942129705.

Lamirel C, Hupé J-M, Lorenceau J. Pupil dynamics during bistable form/motion binding. *J Vision*. 2008;8(6):253, 253a.

Lamme VA, Roelfsema PR. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*. 2000 Nov;23(11):571-9. doi:10.1016/s0166-2236(00)01657-x.

Latawiec D, Martin KA, Meskenaite V. Termination of the geniculocortical projection in the striate cortex of macaque monkey: a quantitative immunoelectron microscopic study. *J Comp Neurol*. 2000 Apr 10;419(3):306-19. doi:10.1002/(sici)1096-9861(20000410)419:3<306::aid-cne4>3.0.co;2-2.

Lawrence SJD, van Mourik T, Kok P, Koopmans PJ, Norris DG, de Lange FP. Laminar Organization of Working Memory Signals in Human Visual Cortex. *Curr Biol*. 2018 Nov 5;28(21):3435-3440.e4. doi:10.1016/j.cub.2018.08.043.

Ledgeway T, Smith AT. Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision. *Vision Res*. 1994 Oct;34(20):2727-40. doi:10.1016/0042-6989(94)90229-1.

Leigh RJ & David SZ. *The Neurology of Eye Movements*, 5 edn, Contemporary Neurology Series (New York, 2015; online edn, Oxford Academic, 1 June 2015). doi:10.1093/med/9780199969289.001.0001.

Leopold DA, Logothetis NK. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*. 1996 Feb 8;379(6565):549-53. doi:10.1038/379549a0.

Leopold DA, Logothetis NK. Multistable phenomena: changing views in perception. *Trends Cogn Sci*. 1999 Jul;3(7):254-264. doi:10.1016/s1364-6613(99)01332-7.

Levelt WJM. *On binocular rivalry*. The Hague, Paris, Mouton 1968.

Li M, Song XM, Xu T, Hu D, Roe AW, Li CY. Subdomains within orientation columns of primary visual cortex. *Sci Adv*. 2019 Jun 5;5(6):eaaw0807. doi:10.1126/sciadv.aaw0807.

Lisberger SG, Morris EJ, Tychsen L. Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annu Rev Neurosci*. 1987;10:97-129. doi:10.1146/annurev.ne.10.030187.000525.

Liu Y, Li M, Zhang X, Lu Y, Gong H, Yin J, Chen Z, Qian L, Yang Y, Andolina IM, Shipp S, Mcloughlin N, Tang S, Wang W. Hierarchical Representation for Chromatic Processing across Macaque V1, V2, and V4. *Neuron*. 2020 Nov 11;108(3):538-550.e5. doi:10.1016/j.neuron.2020.07.037.

Livingstone MS, Hubel DH. Anatomy and physiology of a color system in the primate visual cortex. *J Neurosci*. 1984 Jan;4(1):309-56. doi:10.1523/JNEUROSCI.04-01-00309.1984.

Logothetis NK, Schall JD. Neuronal correlates of subjective visual perception. *Science*. 1989 Aug 18;245(4919):761-3. doi:10.1126/science.2772635.

Logothetis NK, Leopold DA, Sheinberg DL. What is rivalling during binocular rivalry? *Nature*. 1996 Apr 18;380(6575):621-4. doi:10.1038/380621a0.

Logothetis NK. Vision: a window on consciousness. *Sci Am*. 1999 Nov;281(5):69-75.

Lorenceau J, Alais D. Form constraints in motion binding. *Nat Neurosci*. 2001 Jul;4(7):745-51. doi:10.1038/89543.

Lu ZL, Sperling G. Attention-generated apparent motion. *Nature*. 1995 Sep 21;377(6546):237-9. doi:10.1038/377237a0.

Lu ZL, Sperling G. Second-order reversed phi. *Percept Psychophys*. 1999 Aug;61(6):1075-88. doi:10.3758/bf03207615.

Lu ZL, Sperling G. Three-systems theory of human visual motion perception: review and update. *J Opt Soc Am A Opt Image Sci Vis*. 2001 Sep;18(9):2331-70. doi:10.1364/josaa.18.002331. Erratum in: *J Opt Soc Am A Opt Image Sci Vis* 2002 Feb;19(2):413.

Lund JS, Lund RD, Hendrickson AE, Bunt AH, Fuchs AF. The origin of efferent pathways from the primary visual cortex, area 17, of the macaque monkey as shown by retrograde transport of horseradish peroxidase. *J Comp Neurol*. 1975 Dec 1;164(3):287-303. doi:10.1002/cne.901640303.

Lupyan G. Cognitive Penetrability of Perception in the Age of Prediction: Predictive Systems are Penetrable Systems. *Rev.Phil.Psych*. 6, 547–569. 2015. doi:10.1007/s13164-015-0253-4

Lupyan G. Changing What You See by Changing What You Know: The Role of Attention. *Front Psychol*. 2017 May 1;8:553. doi:10.3389/fpsyg.2017.00553.

Maier A, Adams GK, Aura C, Leopold DA. Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. *Front Syst Neurosci*. 2010 Aug 10;4:31. doi:10.3389/fnsys.2010.00031.

Malonek D, Tootell RB, Grinvald A. Optical imaging reveals the functional architecture of neurons processing shape and motion in owl monkey area MT. *Proc Biol Sci*. 1994 Nov 22;258(1352):109-19. doi:10.1098/rspb.1994.0150.

Marić M, Domijan D. Dual counterstream architecture may support separation between vision and predictions. *Conscious Cogn.* 2022 Aug;103:103375. doi:10.1016/j.concog.2022.103375.

Markov NT, Misery P, Falchier A, Lamy C, Vezoli J, Quilodran R, Gariel MA, Giroud P, Ercsey-Ravasz M, Pilaz LJ, Huissoud C, Barone P, Dehay C, Toroczkai Z, Van Essen DC, Kennedy H, Knoblauch K. Weight consistency specifies regularities of macaque cortical networks. *Cereb Cortex.* 2011 Jun;21(6):1254-72. doi:10.1093/cercor/bhq201.

Markov NT, Ercsey-Ravasz M, Lamy C, Ribeiro Gomes AR, Magrou L, Misery P, Giroud P, Barone P, Dehay C, Toroczkai Z, Knoblauch K, Van Essen DC, Kennedy H. The role of long-range connections on the specificity of the macaque inter-areal cortical network. *Proc Natl Acad Sci U S A.* 2013a Mar 26;110(13):5187-92. doi:10.1073/pnas.1218972110. Erratum in: *Proc Natl Acad Sci U S A.* 2013 Oct 15;110(42):1761.

Markov NT, Kennedy H. The importance of being hierarchical. *Curr Opin Neurobiol.* 2013b Apr;23(2):187-94. doi:10.1016/j.conb.2012.12.008.

Markov NT, Ercsey-Ravasz M, Van Essen DC, Knoblauch K, Toroczkai Z, Kennedy H. Cortical high-density counterstream architectures. *Science.* 2013c Nov 1;342(6158):1238406. doi:10.1126/science.1238406.

Markov NT, Ercsey-Ravasz MM, Ribeiro Gomes AR, Lamy C, Magrou L, Vezoli J, Misery P, Falchier A, Quilodran R, Gariel MA, Sallet J, Gamanut R, Huissoud C, Clavagnier S, Giroud P, Sappey-Marinié D, Barone P, Dehay C, Toroczkai Z, Knoblauch K, Van Essen DC, Kennedy H. A weighted and directed inter-areal connectivity matrix for macaque cerebral cortex. *Cereb Cortex.* 2014a Jan;24(1):17-36. doi:10.1093/cercor/bhs270.

Markov NT, Vezoli J, Chameau P, Falchier A, Quilodran R, Huissoud C, Lamy C, Misery P, Giroud P, Ullman S, Barone P, Dehay C, Knoblauch K, Kennedy H. Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J Comp Neurol.* 2014b Jan 1;522(1):225-59. doi:10.1002/cne.23458.

Martinez-Millán L, Holländer H. Cortico-cortical projections from striate cortex of the squirrel monkey (*Saimiri sciureus*). A radioautographic study. *Brain Res.* 1975 Jan 17;83(3):405-17. doi:10.1016/0006-8993(75)90833-1.

Metelli F. The perception of transparency. *Sci Am.* 1974 Apr;230(4):90-8. doi:10.1038/scientificamerican0474-90.

Meyer CH, Lasker AG, Robinson DA. The upper limit of human smooth pursuit velocity. *Vision Res.* 1985;25(4):561-3. doi:10.1016/0042-6989(85)90160-9.

Michalareas G, Vezoli J, van Pelt S, Schoffelen JM, Kennedy H, Fries P. Alpha-Beta and Gamma Rhythms Subserve Feedback and Feedforward Influences among Human Visual Cortical Areas. *Neuron.* 2016 Jan 20;89(2):384-97. doi:10.1016/j.neuron.2015.12.018.

Miles FA, Kawano K, Optican LM. Short-latency ocular following responses of monkey. I. Dependence on temporospatial properties of visual input. *J Neurophysiol.* 1986 Nov;56(5):1321-54. doi:10.1152/jn.1986.56.5.1321.

- Misaki M, Luh WM, Bandettini PA. The effect of spatial smoothing on fMRI decoding of columnar-level organization with linear support vector machine. *J Neurosci Methods*. 2013 Jan 30;212(2):355-61. doi:10.1016/j.jneumeth.2012.11.004.
- Montagnini A, Mamassian P, Perrinet L, Castet E, Masson GS. Bayesian modeling of dynamic motion integration. *J Physiol Paris*. 2007 Jan-May;101(1-3):64-77. doi:10.1016/j.jphysparis.2007.10.013.
- Moreno-Bote R, Rinzel J, Rubin N. Noise-induced alternations in an attractor network model of perceptual bistability. *J Neurophysiol*. 2007 Sep;98(3):1125-39. doi:10.1152/jn.00116.2007.
- Moreno-Bote R, Shpiro A, Rinzel J, Rubin N. Bi-stable depth ordering of superimposed moving gratings. *J Vis*. 2008 Jul 31;8(7):20.1-13. doi:10.1167/8.7.20.
- Moreno-Bote R, Shpiro A, Rinzel J, Rubin N. Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *J Vis*. 2010 Sep 1;10(11):1. doi:10.1167/10.11.1.
- Movshon JA, Adelson EH, Gizzi MS, Newsome WT. The analysis of moving visual patterns. In: *Pattern Recognition Mechanisms: Proceedings of a Study Week/Organized by The Pontifical Academy of Sciences, Casina Pius IV, Vatican City*, edited by Chagas C, Gattass R, Gross C. Berlin: Springer, 1985.
- Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E. Contextual Feedback to Superficial Layers of V1. *Curr Biol*. 2015 Oct 19;25(20):2690-5. doi:10.1016/j.cub.2015.08.057.
- Mumford D. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern*. 1992;66(3):241-51. doi:10.1007/BF00198477.
- Mumford JA, Turner BO, Ashby FG, Poldrack RA. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*. 2012 Feb 1;59(3):2636-43. doi: 10.1016/j.neuroimage.2011.08.076.
- Mumford JA, Davis T, Poldrack RA. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *Neuroimage*. 2014 Dec;103:130-138. doi:10.1016/j.neuroimage.2014.09.026.
- Naber M, Frässle S, Einhäuser W. Perceptual rivalry: reflexes reveal the gradual nature of visual awareness. *PLoS One*. 2011;6(6):e20910. doi:10.1371/journal.pone.0020910.
- Necker LA. "Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid". *London and Edinburgh Philosophical Magazine and Journal of Science*. 1832; 1 (5): 329–337. doi:10.1080/14786443208647909.

Newsome WT, Paré EB. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J Neurosci*. 1988 Jun;8(6):2201-11. doi:10.1523/JNEUROSCI.08-06-02201.1988.

O'Callaghan C, Kveraga K, Shine JM, Adams RB Jr, Bar M. Predictions penetrate perception: Converging insights from brain, behaviour and disorder. *Conscious Cogn*. 2017 Jan;47:63-74. doi:10.1016/j.concog.2016.05.003.

Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*. 1996 Jun 13;381(6583):607-9. doi:10.1038/381607a0.

Olshausen BA: Principles of image representation in visual cortex. In *The Visual Neurosciences*. Edited by Chalupa LM, Werner JS. Boston MA: MIT Press; 2003:1603-1615.

Olshausen BA, Field DJ. Sparse coding of sensory inputs. *Curr Opin Neurobiol*. 2004 Aug;14(4):481-7. doi:10.1016/j.conb.2004.07.007.

Orban GA, Kennedy H. The influence of eccentricity on receptive field types and orientation selectivity in areas 17 and 18 of the cat. *Brain Res*. 1981 Mar 9;208(1):203-8. doi:10.1016/0006-8993(81)90633-8.

Pack CC, Berezovskii VK, Born RT. Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature*. 2001 Dec 20-27;414(6866):905-8. doi:10.1038/414905a.

Pack CC, Livingstone MS, Duffy KR, Born RT. End-stopping and the aperture problem: two-dimensional motion signals in macaque V1. *Neuron*. 2003 Aug 14;39(4):671-80. doi:10.1016/s0896-6273(03)00439-2.

Palmer SE, Marre O, Berry MJ 2nd, Bialek W. Predictive information in a sensory population. *Proc Natl Acad Sci U S A*. 2015 Jun 2;112(22):6908-13. doi:10.1073/pnas.1506855112.

Papale P, Wang F, Morgan T, Chen X, Gilhuis A, Petro LS, Muckli L, Roelfsema PR, Self MW. Feedback brings scene information to the representation of occluded image regions in area V1 of monkeys and humans. *bioRxiv* 2022.11.21.517305; doi:10.1101/2022.11.21.517305.

Pascual-Leone A, Walsh V. Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*. 2001 Apr 20;292(5516):510-2. doi:10.1126/science.1057099.

Passingham RE, Stephan KE, Kötter R. The anatomical basis of functional localization in the cortex. *Nat Rev Neurosci*. 2002 Aug;3(8):606-16. doi:10.1038/nrn893.

Peters JC, Goebel R, Goffaux V. From coarse to fine: Interactive feature processing precedes local feature analysis in human face perception. *Biol Psychol*. 2018 Oct;138:1-10. doi:10.1016/j.biopsycho.2018.07.009.

Petro LS, Vizioli L, Muckli L. Contributions of cortical feedback to sensory processing in primary visual cortex. *Front Psychol.* 2014 Nov 6;5:1223. doi:10.3389/fpsyg.2014.01223.

Pinna B, Werner JS, Spillmann L. The watercolor effect: a new principle of grouping and figure-ground organization. *Vision Res.* 2003 Jan;43(1):43-52. doi:10.1016/s0042-6989(02)00132-3.

Poline JB, Brett M. The general linear model and fMRI: does love last forever? *Neuroimage.* 2012 Aug 15;62(2):871-80. doi:10.1016/j.neuroimage.2012.01.133.

Poort J, Raudies F, Wannig A, Lamme VA, Neumann H, Roelfsema PR. The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron.* 2012 Jul 12;75(1):143-56. doi:10.1016/j.neuron.2012.04.032.

Poort J, Self MW, van Vugt B, Malkki H, Roelfsema PR. Texture Segregation Causes Early Figure Enhancement and Later Ground Suppression in Areas V1 and V4 of Visual Cortex. *Cereb Cortex.* 2016 Oct;26(10):3964-76. doi:10.1093/cercor/bhw235.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012a). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3), 2142-2154. doi:10.1016/j.neuroimage.2011.10.018.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012b). Steps toward optimizing motion artifact removal in functional connectivity MRI; a reply to Carp. *NeuroImage*. doi:10.1016/j.neuroimage.2012.03.017.

Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage.* 2014 Jan 1;84:320-41. doi:10.1016/j.neuroimage.2013.08.048.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. (2007). *Numerical recipes: the art of scientific computing*, 3rd Edition.

Previc FH. Functional specialization in the lower and upper visual fields in humans: its ecological origins and neurophysiological implications. *Behav. Brain Sci.* 1990; 13 519–542. doi:10.1017/S0140525X00080018.

Raftopoulos A. *Cognition and perception: How do psychology and neural science inform philosophy?* MIT Press, Cambridge, MA. 2009.

Raftopoulos, A. The cognitive impenetrability of the content of early vision is a necessary and sufficient condition for purely nonconceptual content. *Philosophical Psychology*, 27 (5) (2014), pp. 601-620.

Raftopoulos, Athanassios, and John Zeimbekis, 'The Cognitive Penetrability of Perception: An Overview', in John Zeimbekis, and Athanassios Raftopoulos, *The Cognitive Penetrability of Perception: New Philosophical Perspectives*. Oxford, 2015; online edn, Oxford Academic, 20 Aug 2015, doi:10.1093/acprof:oso/9780198738916.003.0001.

- Raftopoulos A. Cognitive penetrability and the epistemic role of perception. Cham, Switzerland: Palgrave Macmillan. 2019. doi:10.1007/978-3-030-10445-0.
- Rahnev D, Lau H, de Lange FP. Prior expectation modulates the interaction between sensory and prefrontal regions in the human brain. *J Neurosci*. 2011 Jul 20;31(29):10741-8. doi:10.1523/JNEUROSCI.1478-11.2011.
- Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*. 1999 Jan;2(1):79-87. doi:10.1038/4580.
- Rockland KS, Pandya DN. Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res*. 1979 Dec 21;179(1):3-20. doi:10.1016/0006-8993(79)90485-2.
- Rockland KS, Pandya DN. Cortical connections of the occipital lobe in the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temporal sulcus. *Brain Res*. 1981 May 18;212(2):249-70. doi: 10.1016/0006-8993(81)90461-3.
- Ramachandran VS, Rao VM, Vidyasagar TR. The role of contours in stereopsis. *Nature*. 1973 Apr 6;242(5397):412-4. doi:10.1038/242412a0.
- Ramcharan EJ, Gnadt JW, Sherman SM. The effects of saccadic eye movements on the activity of geniculate relay neurons in the monkey. *Vis Neurosci*. 2001 Mar-Apr;18(2):253-8. doi:10.1017/s0952523801182106.
- Reichardt, W. (1957). Autokorrelations-Auswertung als Funktionsprinzip des Zentralnervensystems. *Zeitschrift für Naturforschung B*, 12, 448 - 457.
- Reppas JB, Usrey WM, Reid RC. Saccadic eye movements modulate visual responses in the lateral geniculate nucleus. *Neuron*. 2002 Aug 29;35(5):961-74. doi:10.1016/s0896-6273(02)00823-1.
- Rockland KS, Pandya DN. Cortical connections of the occipital lobe in the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temporal sulcus. *Brain Res*. 1981 May 18;212(2):249-70. doi:10.1016/0006-8993(81)90461-3.
- Rodman HR, Albright TD. Coding of visual stimulus velocity in area MT of the macaque. *Vision Res* 27: 2035–2048, 1987. doi:10.1016/0042-6989(87)90118-0.
- Rolls ET, Tovee MJ. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol*. 1995 Feb;73(2):713-26. doi:10.1152/jn.1995.73.2.713.
- Roy CS, Sherrington CS. On the Regulation of the Blood-supply of the Brain. *J Physiol*. 1890 Jan;11(1-2):85-158.17. doi:10.1113/jphysiol.1890.sp000321.
- Rubin E. The Danish-language Synsoplevede Figurer ("Visual Figures"). Gyldendal 1915, Vol. 1-2.

Rubin N, Hupé J-M. Dynamics of perceptual bistability: Plaids and binocular rivalry compared. Edited by Alais, D., Blake, R. *Binocular Rivalry*, MIT Press, 2005; 8:137-154.

Sahraie A, Barbur JL. Pupil response triggered by the onset of coherent motion. *Graefes Arch Clin Exp Ophthalmol*. 1997 Aug;235(8):494-500. doi:10.1007/BF00947006.

Salin PA, Girard P, Kennedy H, Bullier J. Visuotopic organization of corticocortical connections in the visual system of the cat. *J Comp Neurol*. 1992 Jun 22;320(4):415-34. doi:10.1002/cne.903200402.

Sato T, Uchida G, Tanifuji M. Cortical columnar organization is reconsidered in inferior temporal cortex. *Cereb Cortex* 2009 19:1870–1888, doi:10.1093/cercor/bhn218.

Savitzky A & Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 1964 36 (8), 1627-1639. doi:10.1021/ac60214a047.

Schneider M, Kemper VG, Emmerling TC, De Martino F, Goebel R. Columnar clusters in the human motion complex reflect consciously perceived motion axis. *Proc Natl Acad Sci U S A*. 2019 Mar 12;116(11):5096-5101. doi:10.1073/pnas.1814504116.

Self MW, van Kerkoerle T, Supèr H, Roelfsema PR. Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Curr Biol*. 2013 Nov 4;23(21):2121-9. doi:10.1016/j.cub.2013.09.013.

Self MW, van Kerkoerle T, Goebel R, Roelfsema PR. Benchmarking laminar fMRI: Neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. *Neuroimage*. 2019 Aug 15;197:806-817. doi:10.1016/j.neuroimage.2017.06.045.

Sheliga BM, Quaia C, FitzGibbon EJ, Cumming BG. Anisotropy in spatial summation properties of human Ocular-Following Response (OFR). *Vision Res*. 2015 Apr;109(Pt A):11-9. doi:10.1016/j.visres.2015.02.015.

Shipp S, Zeki S. The Organization of Connections between Areas V5 and V1 in Macaque Monkey Visual Cortex. *Eur J Neurosci*. 1989;1(4):309-32. doi:10.1111/j.1460-9568.1989.tb00798.x.

Shipp S, Adams RA, Friston KJ. Reflections on agranular architecture: predictive coding in the motor cortex. *Trends Neurosci*. 2013 Dec;36(12):706-16. doi:10.1016/j.tins.2013.09.004.

Shipp S. Neural Elements for Predictive Coding. *Front Psychol*. 2016 Nov 18;7:1792. doi:10.3389/fpsyg.2016.01792.

Shapiro A, Curtu R, Rinzel J, Rubin N. Dynamical characteristics common to neuronal competition models. *J Neurophysiol*. 2007 Jan;97(1):462-73. doi:10.1152/jn.00604.2006.

Shapiro A, Moreno-Bote R, Rubin N, Rinzel J. Balance between noise and adaptation in competition models of perceptual bistability. *J Comput Neurosci*. 2009 Aug;27(1):37-54. doi:10.1007/s10827-008-0125-3.

Slotnick SD, Thompson WL, Kosslyn SM. Visual mental imagery induces retinotopically organized activation of early visual areas. *Cereb Cortex*. 2005 Oct;15(10):1570-83. doi:10.1093/cercor/bhi035.

Smith AT. Coherence of plaids comprising components of disparate spatial frequencies. *Vision Res*. 1992 Feb;32(2):393-7. doi:10.1016/0042-6989(92)90148-c.

Smith AT. Correspondence-based and energy-based detection of second-order motion in human vision. *J Opt Soc Am A Opt Image Sci Vis*. 1994 Jul;11(7):1940-8. doi:10.1364/josaa.11.001940.

Snowden RJ, Hammett ST. Subtractive and divisive adaptation in the human visual system. *Nature*. 1992 Jan 16;355(6357):248-50. doi:10.1038/355248a0.

Sousa T, Duarte JV, Costa GN, Kemper VG, Martins R, Goebel R, Castelo-Branco M. The dual nature of the BOLD signal: Responses in visual area hMT+ reflect both input properties and perceptual decision. *Hum Brain Mapp*. 2021 Apr 15;42(6):1920-1929. doi:10.1002/hbm.25339.

Spatz WB, Tigges J, Tigges M. Subcortical projections, cortical associations, and some intrinsic interlaminar connections of the striate cortex in the squirrel monkey (*Saimiri*). *J Comp Neurol*. 1970 Oct;140(2):155-74. doi:10.1002/cne.901400203.

Spering M, Gegenfurtner KR. Contextual effects on motion perception and smooth pursuit eye movements. *Brain Res*. 2008 Aug 15;1225:76-85. doi:10.1016/j.brainres.2008.04.061.

Srinivasan MV, Laughlin SB, Dubs A. Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc Lond B Biol Sci*. 1982 Nov 22;216(1205):427-59. doi:10.1098/rspb.1982.0085.

Sterzer P, Haynes JD, Rees G. Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. *Neuroimage*. 2006 Sep;32(3):1308-16. doi:10.1016/j.neuroimage.2006.05.029.

Stone LS, Krauzlis RJ. Shared motion signals for human perceptual decisions and oculomotor actions. *J Vis*. 2003 Dec 2;3(11):725-36. doi:10.1167/3.11.7.

Stoner GR, Albright TD, Ramachandran VS. Transparency and coherence in human motion perception. *Nature*. 1990 Mar 8;344(6262):153-5. doi:10.1038/344153a0.

Stoner GR, Albright TD. Motion coherency rules are form-cue invariant. *Vision Res*. 1992a Mar;32(3):465-75. doi:10.1016/0042-6989(92)90238-e.

Stoner GR, Albright TD. Neural correlates of perceptual motion coherence. *Nature*. 1992b Jul 30;358(6385):412-4. doi:10.1038/358412a0.

Stoner GR, Albright TD. The interpretation of visual motion: evidence for surface segmentation mechanisms. *Vision Res*. 1996 May;36(9):1291-310. doi:10.1016/0042-6989(95)00195-6.

- Stoner GR, Albright TD. Luminance contrast affects motion coherency in plaid patterns by acting as a depth-from-occlusion cue. *Vision Res.* 1998 Feb;38(3):387-401. doi:10.1016/s0042-6989(97)00132-6.
- Strogatz SH. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering.* CRC press, 2018.
- Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egnér T. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci.* 2008 Sep;11(9):1004-6. doi:10.1038/nn.2163.
- Summerfield C, Wyart V, Johnen VM, de Gardelle V. Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. *Front Hum Neurosci.* 2011 Jul 28;5:67. doi:10.3389/fnhum.2011.00067.
- Sun P, Chubb C, Sperling G. A moving-barber-pole illusion. *J Vis.* 2014 May 1;14(5):1. doi:10.1167/14.5.1.
- Sun P, Chubb C, Sperling G. Two mechanisms that determine the Barber-Pole Illusion. *Vision Res.* 2015 Jun;111(Pt A):43-54. doi:10.1016/j.visres.2015.04.002.
- Tanaka K. Inferotemporal cortex and object vision. *Annu Rev Neurosci.* 1996;19:109-39. doi:10.1146/annurev.ne.19.030196.000545.
- Tang S, Lee TS, Li M, Zhang Y, Xu Y, Liu F, Teo B, Jiang H. Complex Pattern Selectivity in Macaque Primary Visual Cortex Revealed by Large-Scale Two-Photon Imaging. *Curr Biol.* 2018a Jan 8;28(1):38-48.e3. doi:10.1016/j.cub.2017.11.039.
- Tang S, Zhang Y, Li Z, Li M, Liu F, Jiang H, Lee TS. Large-scale two-photon imaging revealed super-sparse population codes in the V1 superficial layer of awake monkeys. *Elife.* 2018b Apr 26;7:e33370. doi:10.7554/eLife.33370.
- Tavassoli A, Ringach DL. When your eyes see more than you do. *Curr Biol.* 2010 Feb 9;20(3):R93-4. doi:10.1016/j.cub.2009.11.048.
- Thier P, Ilg UJ. The neural basis of smooth-pursuit eye movements. *Curr Opin Neurobiol.* 2005 Dec;15(6):645-52. doi:10.1016/j.conb.2005.10.013.
- Thomas NA, Elias LJ. Upper and lower visual field differences in perceptual asymmetries. *Brain Res.* 2011 Apr 28;1387:108-15. doi:10.1016/j.brainres.2011.02.063.
- Tigges J, Spatz WB, Tigges M. Reciprocal point-to-point connections between parastriate and striate cortex in the squirrel monkey (*Saimiri*). *J Comp Neurol.* 1973 Apr 15;148(4):481-9. doi:10.1002/cne.901480406.
- Tong F. Primary visual cortex and visual awareness. *Nat Rev Neurosci.* 2003 Mar;4(3):219-29. doi:10.1038/nrn1055.

Tsao DY, Conway BR, Livingstone MS. Receptive fields of disparity-tuned simple cells in macaque V1. *Neuron*. 2003 Apr 10;38(1):103-14. doi:10.1016/s0896-6273(03)00150-8.

Tsao DY, Livingstone MS. Mechanisms of face perception. *Annu Rev Neurosci*. 2008;31:411-37. doi:10.1146/annurev.neuro.30.051606.094238.

Tsodyks M, Kenet T, Grinvald A, Arieli A. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*. 1999 Dec 3;286(5446):1943-6. doi:10.1126/science.286.5446.1943.

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci*. 2001 Aug;4(8):832-8. doi:10.1038/90547.

Tullis T, & Albert B. *Measuring the User Experience*. Elsevier, Amsterdam, 2008.

Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010 Jun;29(6):1310-20. doi:10.1109/TMI.2010.2046908.

Tuten WS, Harmening WM. Foveal vision. *Curr Biol*. 2021 Jun 7;31(11):R701-R703. doi:10.1016/j.cub.2021.03.097.

Tychsen L, Lisberger SG. Visual motion processing for the initiation of smooth-pursuit eye movements in humans. *J Neurophysiol*. 1986 Oct;56(4):953-68. doi:10.1152/jn.1986.56.4.953.

van Hateren JH, Ruderman DL. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc Biol Sci*. 1998 Dec 7;265(1412):2315-20. doi:10.1098/rspb.1998.0577.

van Hateren JH, van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc Biol Sci*. 1998 Mar 7;265(1394):359-66. doi:10.1098/rspb.1998.0303.

van Kemenade BM, Seymour K, Christophel TB, Rothkirch M, Sterzer P. Decoding pattern motion information in V1. *Cortex*. 2014 Aug;57:177-87. doi:10.1016/j.cortex.2014.04.014.

van Kemenade BM, Wilbertz G, Müller A, Sterzer P. Non-stimulated regions in early visual cortex encode the contents of conscious visual perception. *Hum Brain Mapp*. 2022 Mar;43(4):1394-1402. doi:10.1002/hbm.25731.

van Kerkoerle T, Self MW, Roelfsema PR. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat Commun*. 2017 Jan 5;8:13804. doi:10.1038/ncomms13804. Erratum in: *Nat Commun*. 2017 May 03;8:15555.

van Santen JP, Sperling G. Temporal covariance model of human motion perception. *J Opt Soc Am A*. 1984 May;1(5):451-73. doi:10.1364/josaa.1.000451.

- van Santen JP, Sperling G. Elaborated Reichardt detectors. *J Opt Soc Am A*. 1985 Feb;2(2):300-21. doi:10.1364/josaa.2.000300.
- Vezoli J, Magrou L, Goebel R, Wang XJ, Knoblauch K, Vinck M, Kennedy H. Cortical hierarchy, dual counterstream architecture and the importance of top-down generative networks. *Neuroimage*. 2021 Jan 15;225:117479. doi:10.1016/j.neuroimage.2020.117479.
- Vinje WE, Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*. 2000 Feb 18;287(5456):1273-6. doi:10.1126/science.287.5456.1273.
- von Grünau M, Dubé S. Comparing local and remote motion aftereffects. *Spat Vis*. 1992;6(4):303-14. doi:10.1163/156856892x00145.
- von Grünau M, Dubé S. Ambiguous plaids: switching between coherence and transparency. *Spat Vis*. 1993;7(3):199-211. doi:10.1163/156856893x00360.
- Von Helmholtz, H. & Southall, J. P. C. (1967/1924–1925). *Treatise on physiological optics*. Rochester, NY: The Optical Society of America. Originally published as *Handbuch der Physiologischen Optik* in 1867, Leipzig, Germany: Leopold Voss.
- von Schiller P. Stroboskopische alternativversuche. *Psychologische Forschung*, 1933; 17, 179–214.
- Wallach H. *Über visuell wahrgenommene Bewegungsrichtung* *Psychologische Forschung*, 1935; 20:325-380.
- Wallach, H., & O'Connell, D.N. (1953). The kinetic depth effect. *Journal of experimental psychology*, 45 4, 205-17 .
- Wallisch P, Movshon JA. Responses of neurons in macaque MT to unikinetic plaids. *J Neurophysiol*. 2019 Nov 1;122(5):1937-1945. doi:10.1152/jn.00486.2019.
- Wang G, Tanaka K, Tanifuji M. Optical imaging of functional organization in the monkey inferotemporal cortex. *Science*. 1996 Jun 14;272(5268):1665-8. doi:10.1126/science.272.5268.1665.
- Wang G, Tanifuji M, Tanaka K. Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neurosci Res*. 1998 Sep;32(1):33-46. doi:10.1016/s0168-0102(98)00062-5.
- Wang JZ, Kowler E. Micropursuit and the control of attention and eye movements in dynamic environments. *J Vis*. 2021 Aug 2;21(8):6. doi:10.1167/jov.21.8.6.
- Wang M, Hou J, Magrou L, Autio JA, Misery P, Coalson T, Reid E, Xu Y, Lamy C, Falchier A, Zhang Q, Poo M-M, Dehay C, Glasser MF, Hayashi T, Knoblauch K, Van Essen D, Shen Z, Kennedy H. Retinotopic organization of feedback projections in primate early visual cortex: implications for active vision. *bioRxiv* 2022.04.27.489651; doi:10.1101/2022.04.27.489651

- Watanabe K. Optokinetic nystagmus with spontaneous reversal of transparent motion perception. *Exp Brain Res.* 1999 Nov;129(1):156-60. doi:10.1007/s002210050946.
- Watson AB, Ahumada AJ Jr. Model of human visual-motion sensing. *J Opt Soc Am A.* 1985 Feb;2(2):322-41. doi:10.1364/josaa.2.000322. PMID: 3973764.
- Wei M, Sun F. The alternation of optokinetic responses driven by moving stimuli in humans. *Brain Res.* 1998 Dec 7;813(2):406-10. doi:10.1016/s0006-8993(98)01046-4.
- Weilnhammer VA, Ludwig K, Hesselmann G, Sterzer P. Frontoparietal cortex mediates perceptual transitions in bistable perception. *J Neurosci.* 2013 Oct 2;33(40):16009-15. doi:10.1523/JNEUROSCI.1418-13.2013.
- Weilnhammer V, Stuke H, Hesselmann G, Sterzer P, Schmack K. A predictive coding account of bistable perception - a model-based fMRI study. *PLoS Comput Biol.* 2017 May 15;13(5):e1005536. doi:10.1371/journal.pcbi.1005536.
- Weilnhammer V, Fritsch M, Chikermane M, Eckert AL, Kanthak K, Stuke H, Kaminski J, Sterzer P. An active role of inferior frontal cortex in conscious experience. *Curr Biol.* 2021 Jul 12;31(13):2868-2880.e8. doi:10.1016/j.cub.2021.04.043.
- Weiss, Y. and Adelson, E. H. (1998). Slow and smooth: a Bayesian theory for the combination of local motion signals in human vision, *AI Memo 1624/CBCL*, 158, Massachusetts Institute of Technology, Cambridge, MA.
- Wilbertz G, Ketkar M, Guggenmos M, Sterzer P. Combined fMRI- and eye movement-based decoding of bistable plaid motion perception. *Neuroimage.* 2018 May 1;171:190-198. doi:10.1016/j.neuroimage.2017.12.094.
- Wilson HR, Ferrera VP, Yo C. A psychophysically motivated model for two-dimensional motion perception. *Vis Neurosci.* 1992 Jul;9(1):79-97. doi:10.1017/s0952523800006386.
- Wilson HR, Kim J. A model for motion coherence and transparency. *Vis Neurosci.* 1994 Nov-Dec;11(6):1205-20. doi:10.1017/s0952523800007008.
- Wilson HR. *Spikes, decisions, and actions: Dynamical foundations of neuroscience* Oxford University Press, Oxford (1999).
- Wong-Riley MT. Demonstration of geniculocortical and callosal projection neurons in the squirrel monkey by means of retrograde axonal transport of horseradish peroxidase. *Brain Res.* 1974 Oct 18;79(2):267-72. doi:10.1016/0006-8993(74)90415-6.
- Wuerger S, Shapley R, Rubin N. "On the Visually Perceived Direction of Motion" by Hans Wallach: 60 Years Later. *Perception*, 1996; 25(11), 1317–1367. doi:10.1068/p251317.
- Yacoub E, Harel N, Ugurbil K. High-field fMRI unveils orientation columns in humans. *Proc Natl Acad Sci U S A.* 2008 Jul 29;105(30):10607-12. doi:10.1073/pnas.0804110105.

Yo C, Wilson HR. Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Res.* 1992 Jan;32(1):135-47. doi:10.1016/0042-6989(92)90121-x.

Yon D, de Lange FP, Press C. The Predictive Brain as a Stubborn Scientist. *Trends Cogn Sci.* 2019 Jan;23(1):6-8. doi:10.1016/j.tics.2018.10.003.

Yoshioka T, Dow BM. Color, orientation and cytochrome oxidase reactivity in areas V1, V2 and V4 of macaque monkey visual cortex. *Behav Brain Res.* 1996 Apr;76(1-2):71-88. doi:10.1016/0166-4328(95)00184-0.

Zeki S. Area V5-a microcosm of the visual brain. *Front Integr Neurosci.* 2015 Apr 1;9:21. doi:10.3389/fnint.2015.00021.

Zhang X, Xu Q, Jiang Y, Wang Y. The interaction of perceptual biases in bistable perception. *Sci Rep.* 2017 Feb 6;7:42018. doi:10.1038/srep42018.

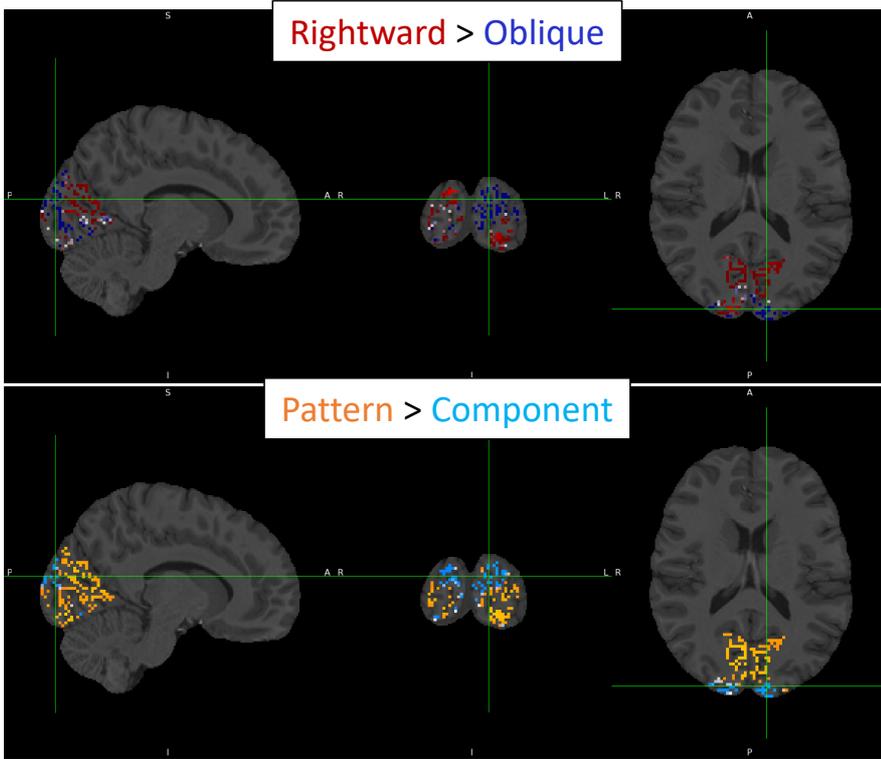
Zimmermann J, Goebel R, De Martino F, van de Moortele PF, Feinberg D, Adriany G, Chaimow D, Shmuel A, Uğurbil K, Yacoub E. Mapping the organization of axis of motion selective features in human area MT using high-field fMRI. *PLoS One.* 2011;6(12):e28716. doi:10.1371/journal.pone.0028716.

Zipser K, Lamme VA, Schiller PH. Contextual modulation in primary visual cortex. *J Neurosci.* 1996 Nov 15;16(22):7376-89. doi:10.1523/JNEUROSCI.16-22-07376.1996.

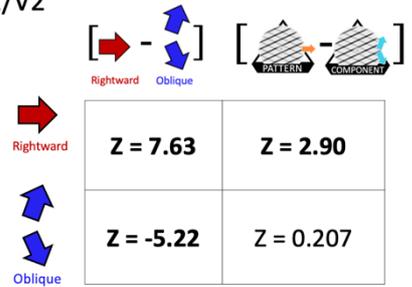
SUPPLEMENTARY

S1

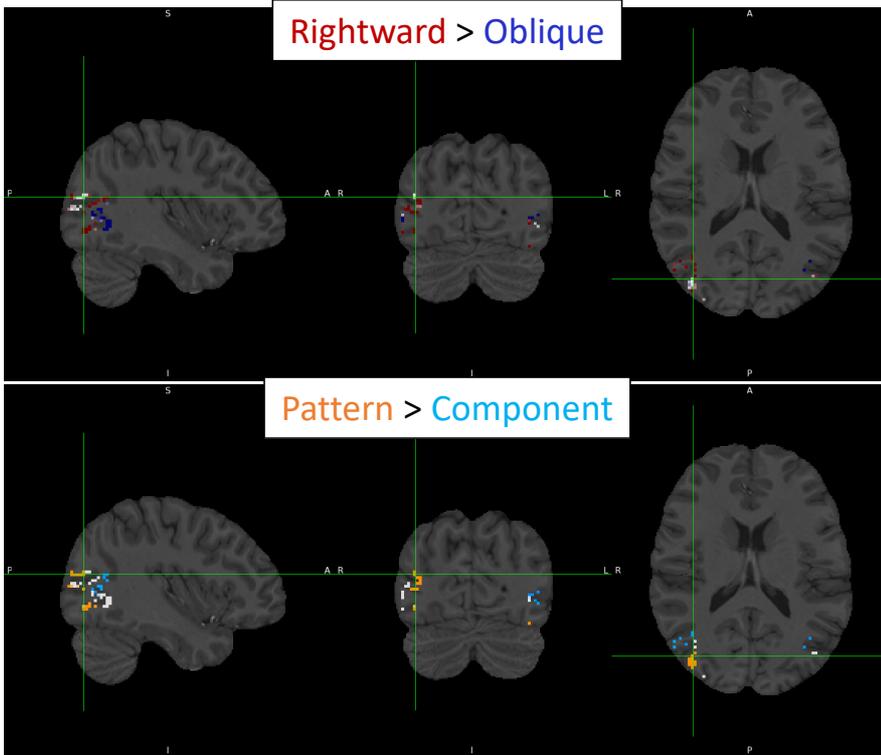
V1/V2



V1/V2



hMT+



hMT+

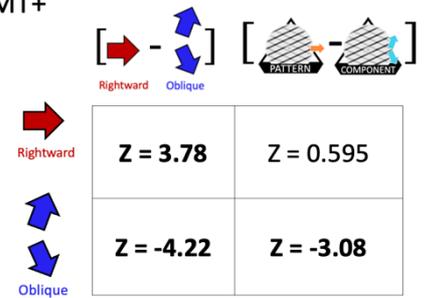


Figure S1.1 : Direction-selectivity and Perceptual decision maps in subject 1

Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).

Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).

Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S2

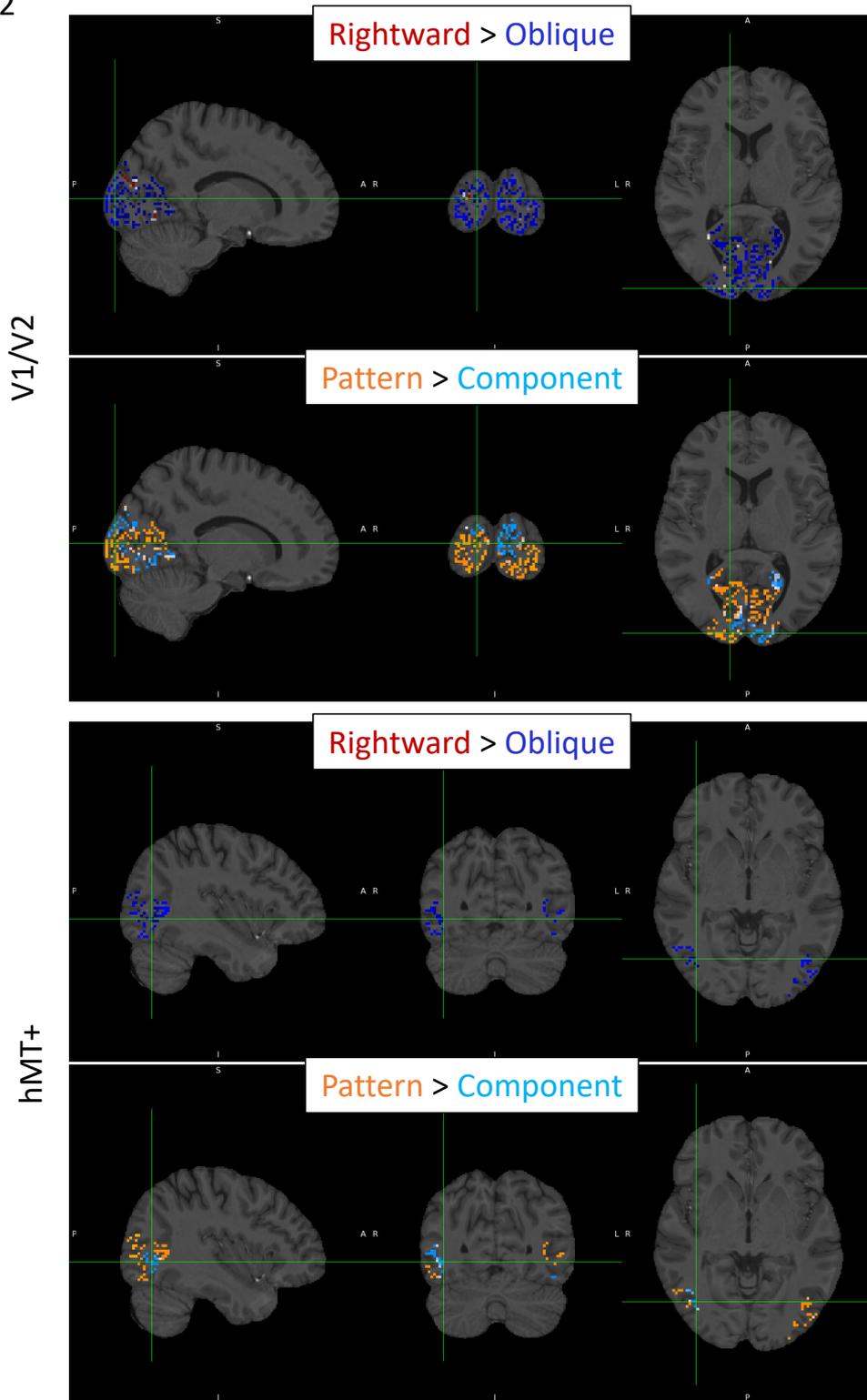


Figure S1.2 : Direction-selectivity and Perceptual decision maps in subject 2

Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Oblique-selective (blue) activity ($p < 0.01$ uncorr) and failed in detecting Rightward selectivity.

Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).

This subject was discarded from the group-level analyses because of the undetected subdomains.

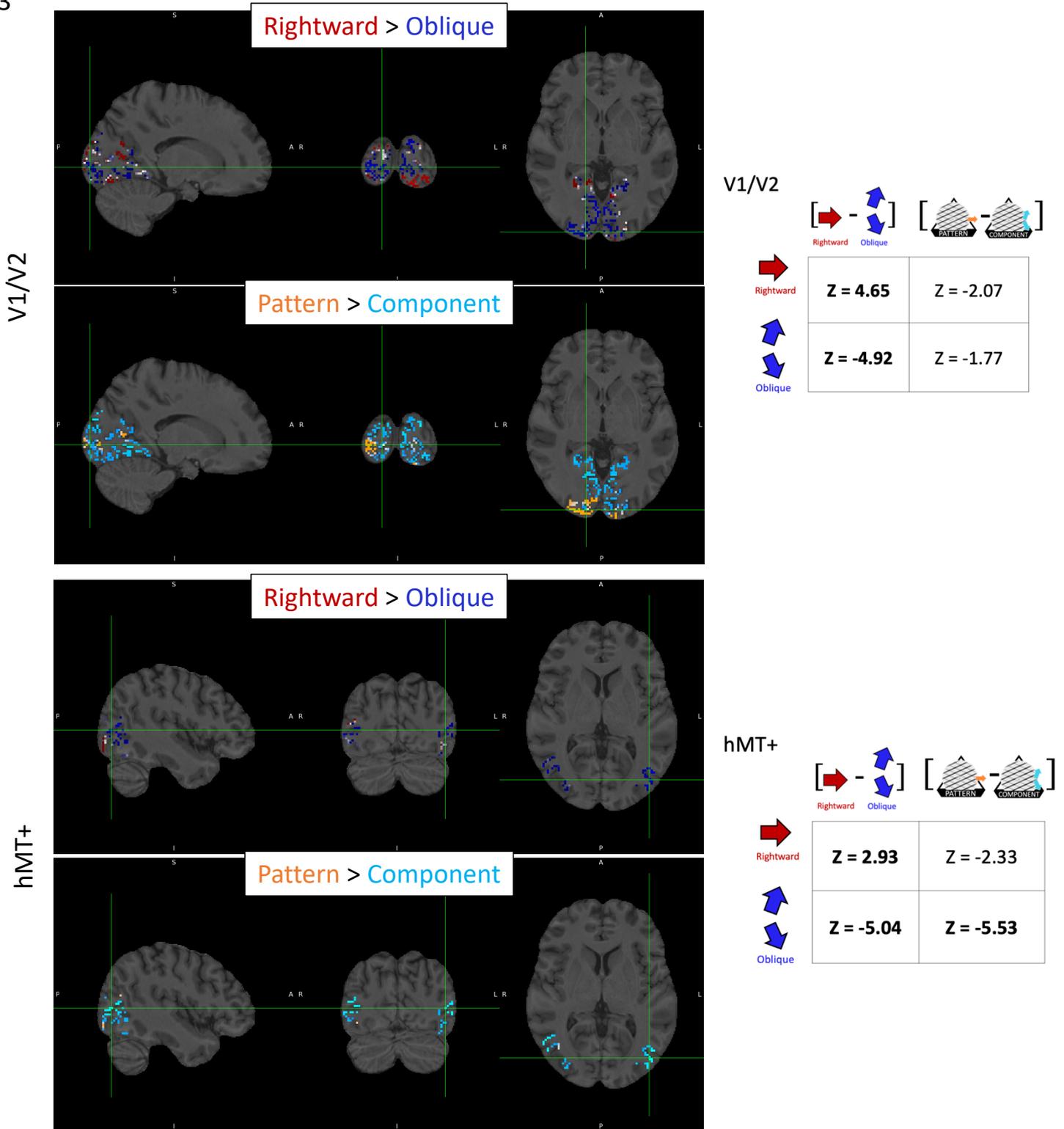


Figure S1.3 : Direction-selectivity and Perceptual decision maps in subject 3
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (in blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

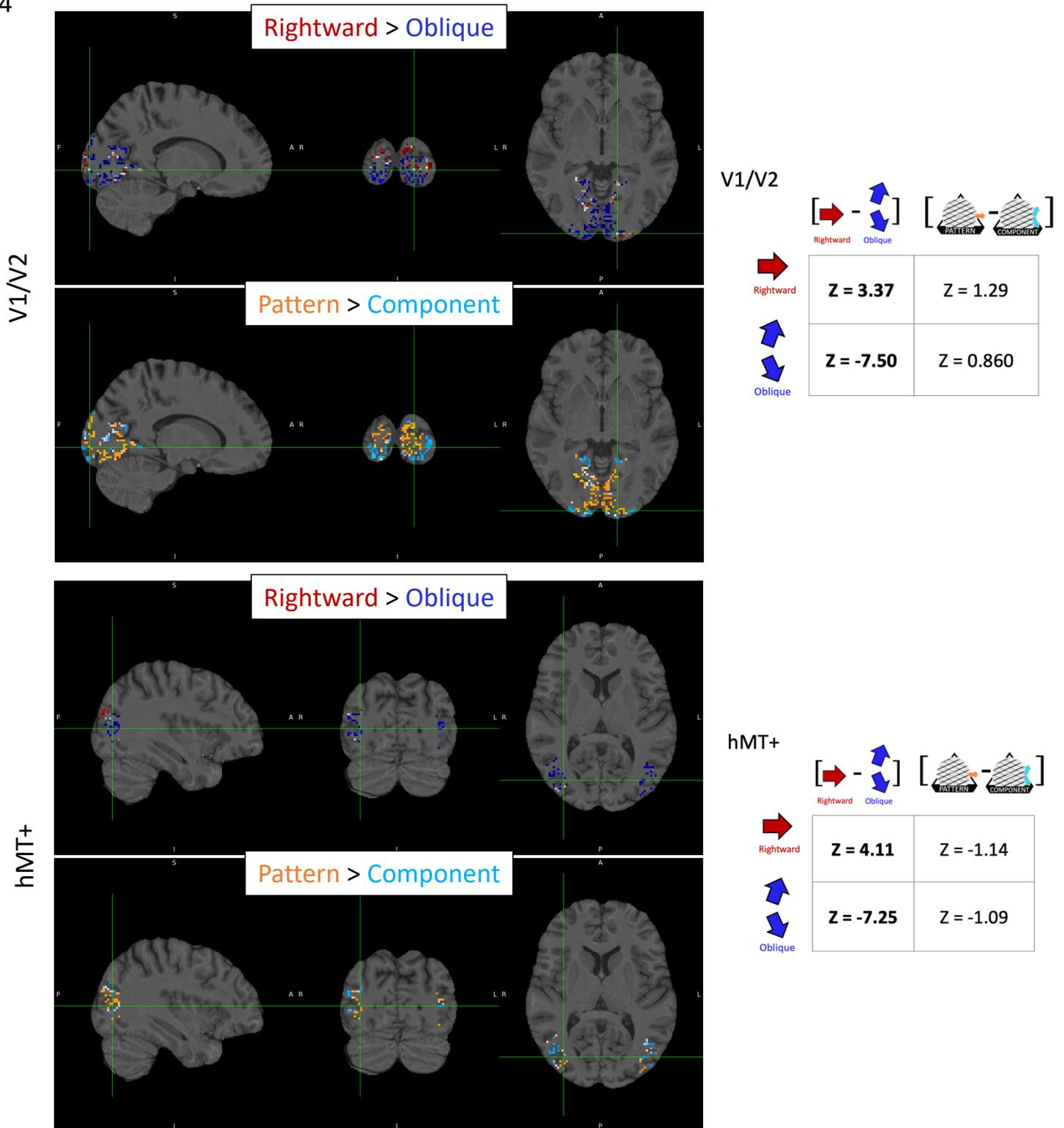
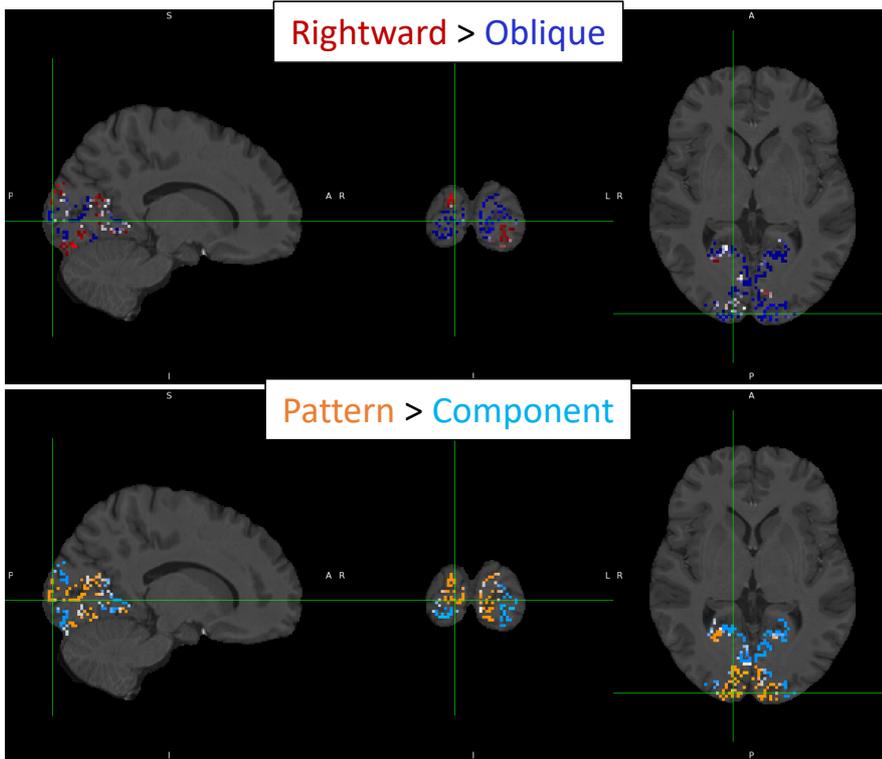


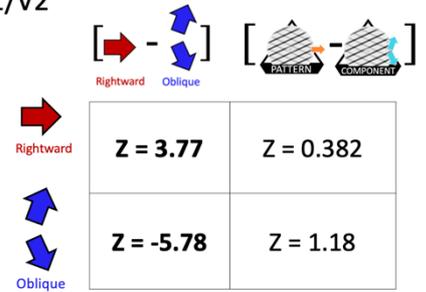
Figure S1.4 : Direction-selectivity and Perceptual decision maps in subject 4
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S5

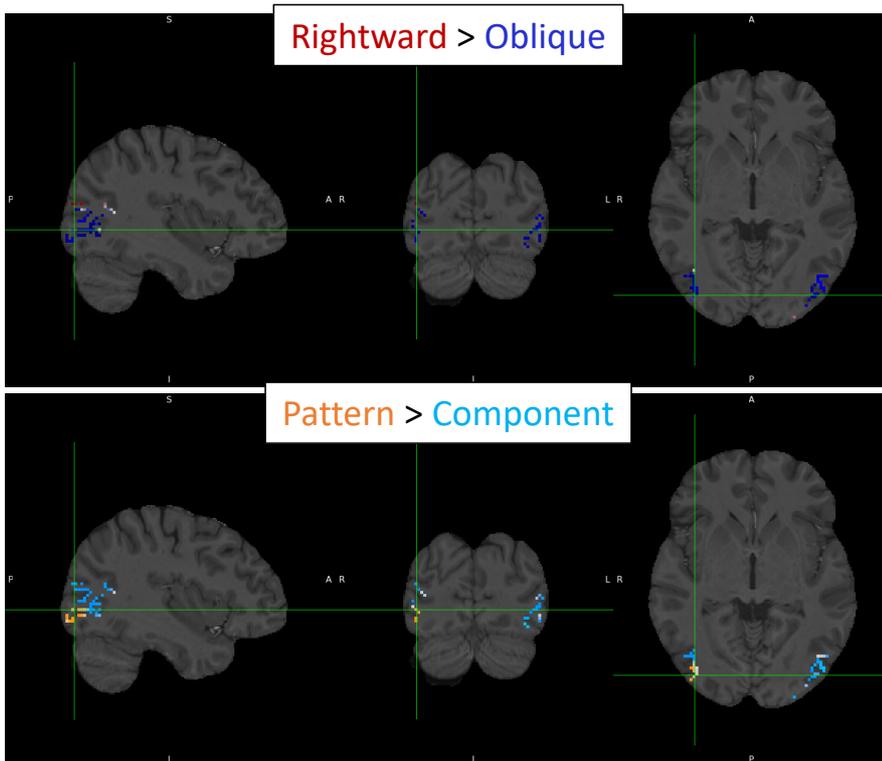
V1/V2



V1/V2



hMT+



hMT+

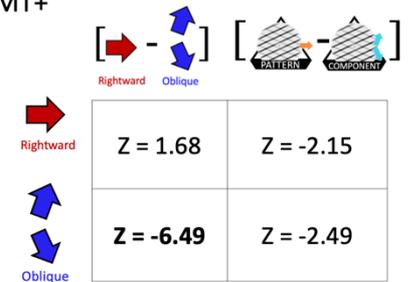


Figure S1.5 : Direction-selectivity and Perceptual decision maps in subject 5
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

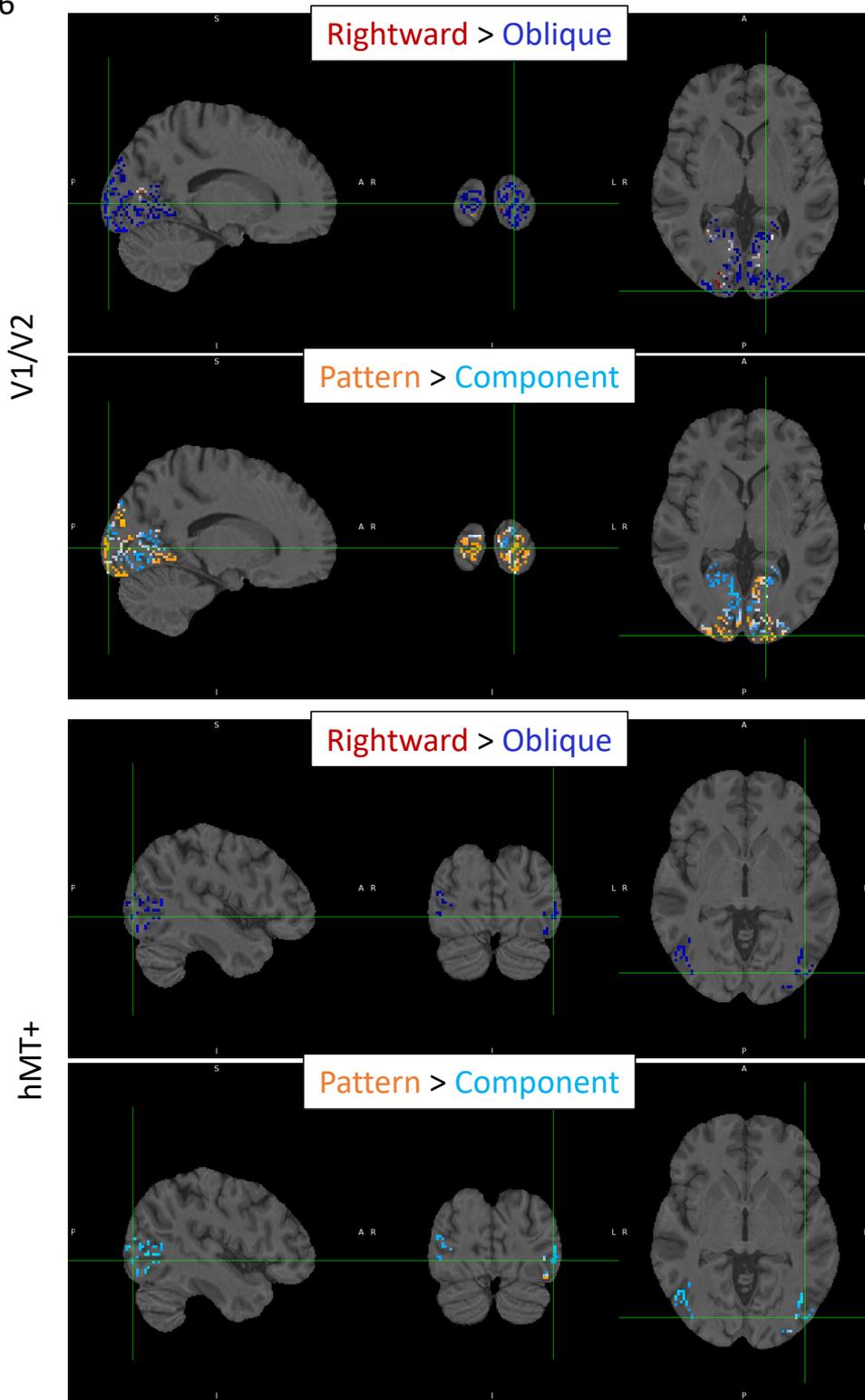


Figure S1.6: Direction-selectivity and Perceptual decision maps in subject 6
Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed predominant response to Oblique directions (blue) ($p < 0.01$ uncorr) and model failed to detect Rightward selectivity.
Bistable motion: Pattern > Component contrast resulted in Component (cyan) and Pattern-selective activity in V1/V2 ($p < 0.01$ uncorr).
This subject was discarded from the group-level analyses because of the undetected subdomains.

S7

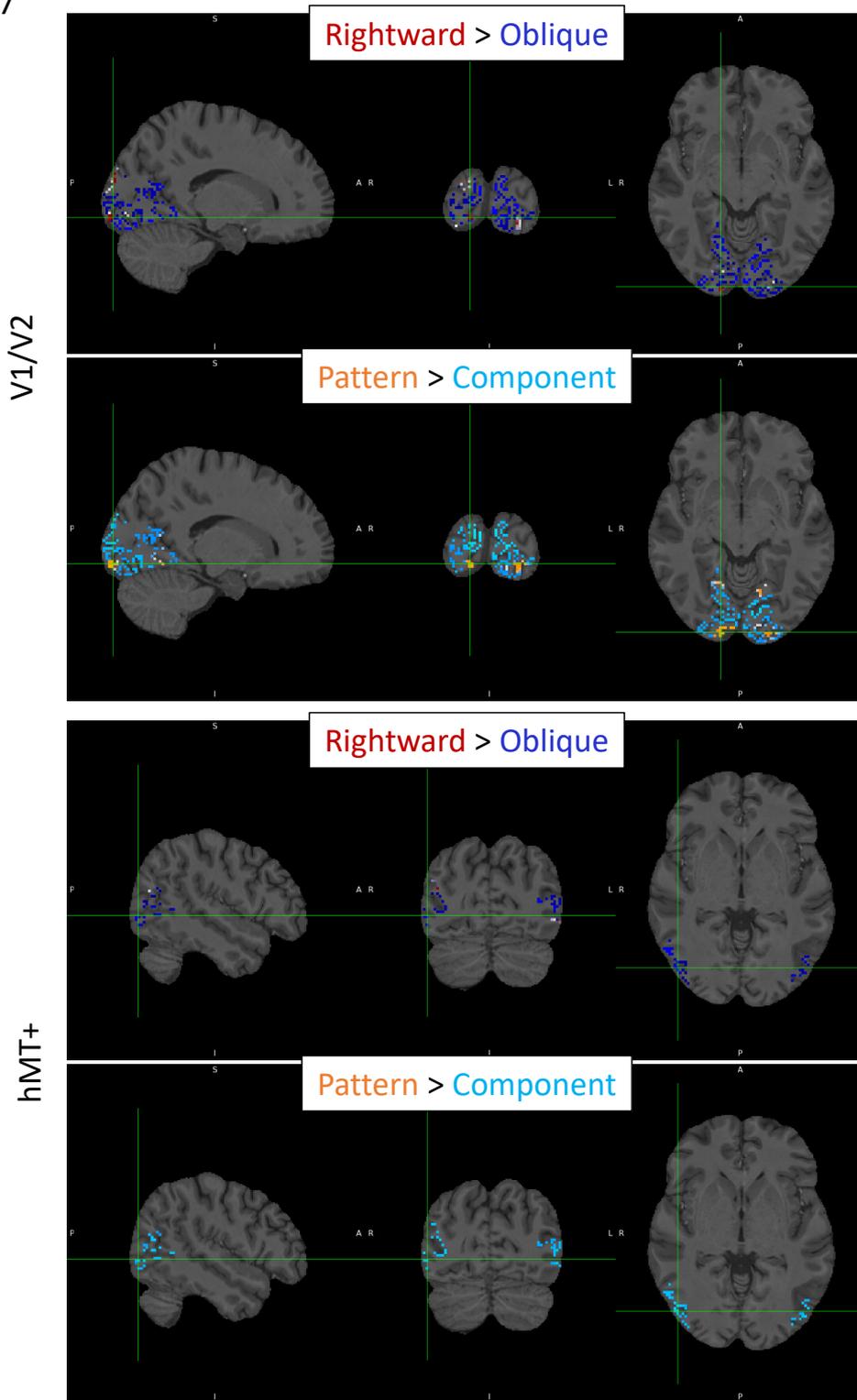


Figure S1.7: Direction-selectivity and Perceptual decision maps in subject 7
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed predominant response to Oblique directions (blue) ($p < 0.01$ uncorr) and model failed to detect Rightward selectivity.
 Bistable motion: Pattern > Component contrast resulted in predominant Component-selective (cyan) response in V1/V2 ($p < 0.01$ uncorr).
 This subject was discarded from the group-level analyses because of the undetected subdomains.

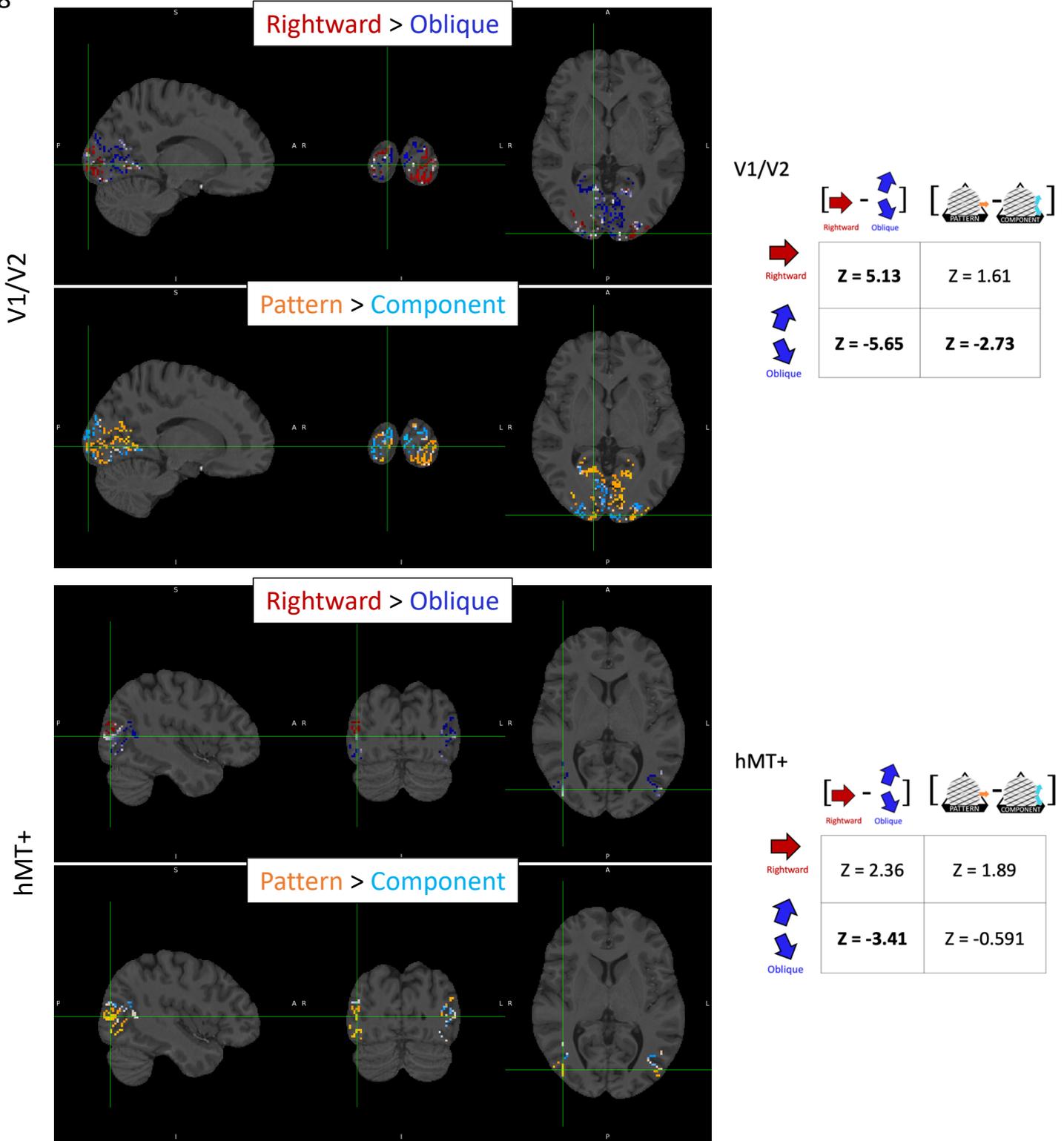


Figure S1.8: Direction-selectivity and Perceptual decision maps in subject 8
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S9

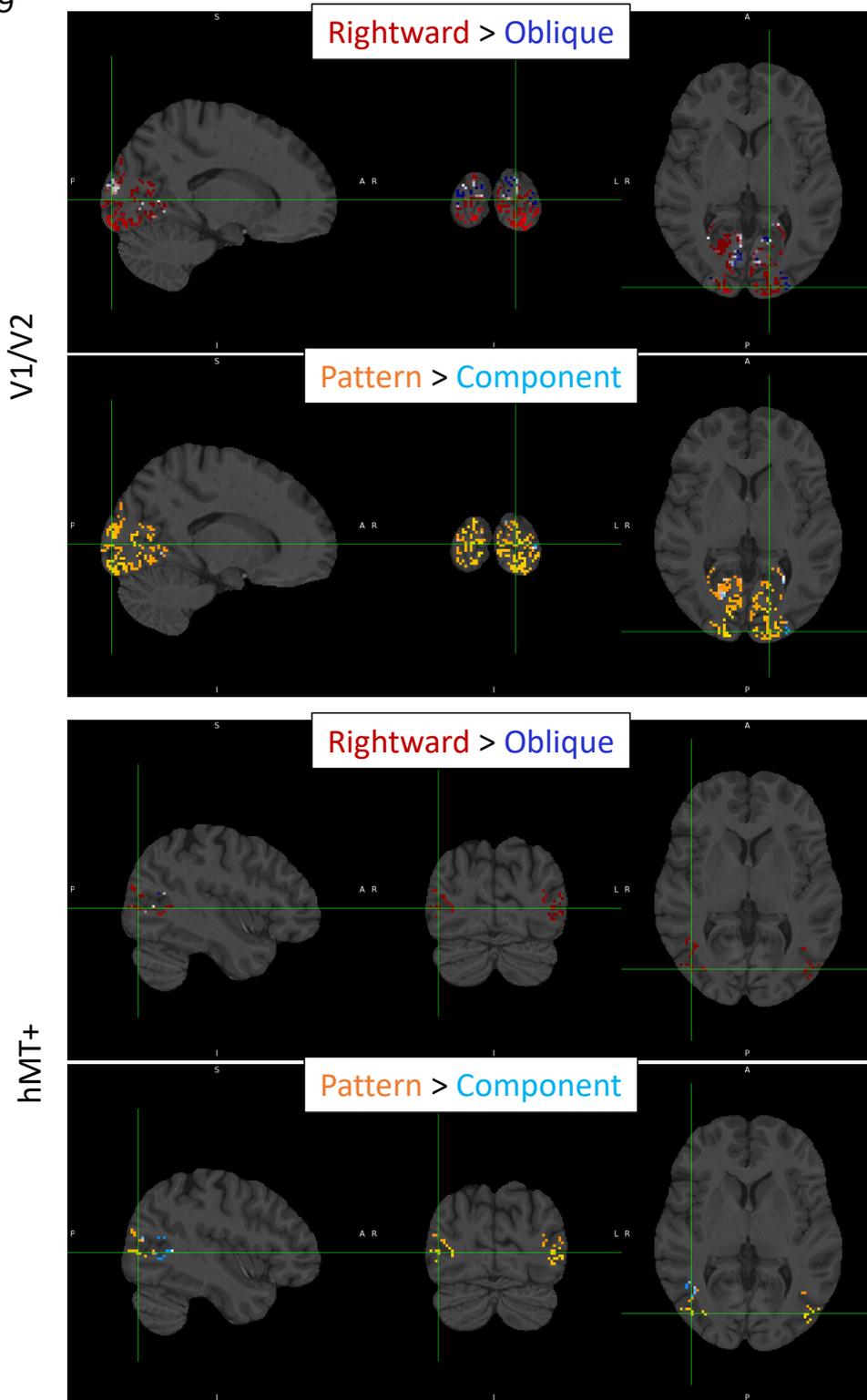


Figure S1.9: Direction-selectivity and Perceptual decision maps in subject 9
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed predominant response to Rightward directions (red) ($p < 0.01$ uncorr) and model failed to detect Oblique selectivity.
 Bistable motion: Overall, Pattern > Component contrast resulted in predominant response to Pattern state (orange) ($p < 0.01$ uncorr).
 This subject was discarded from the group-level analyses because of the undetected subdomains.

S10

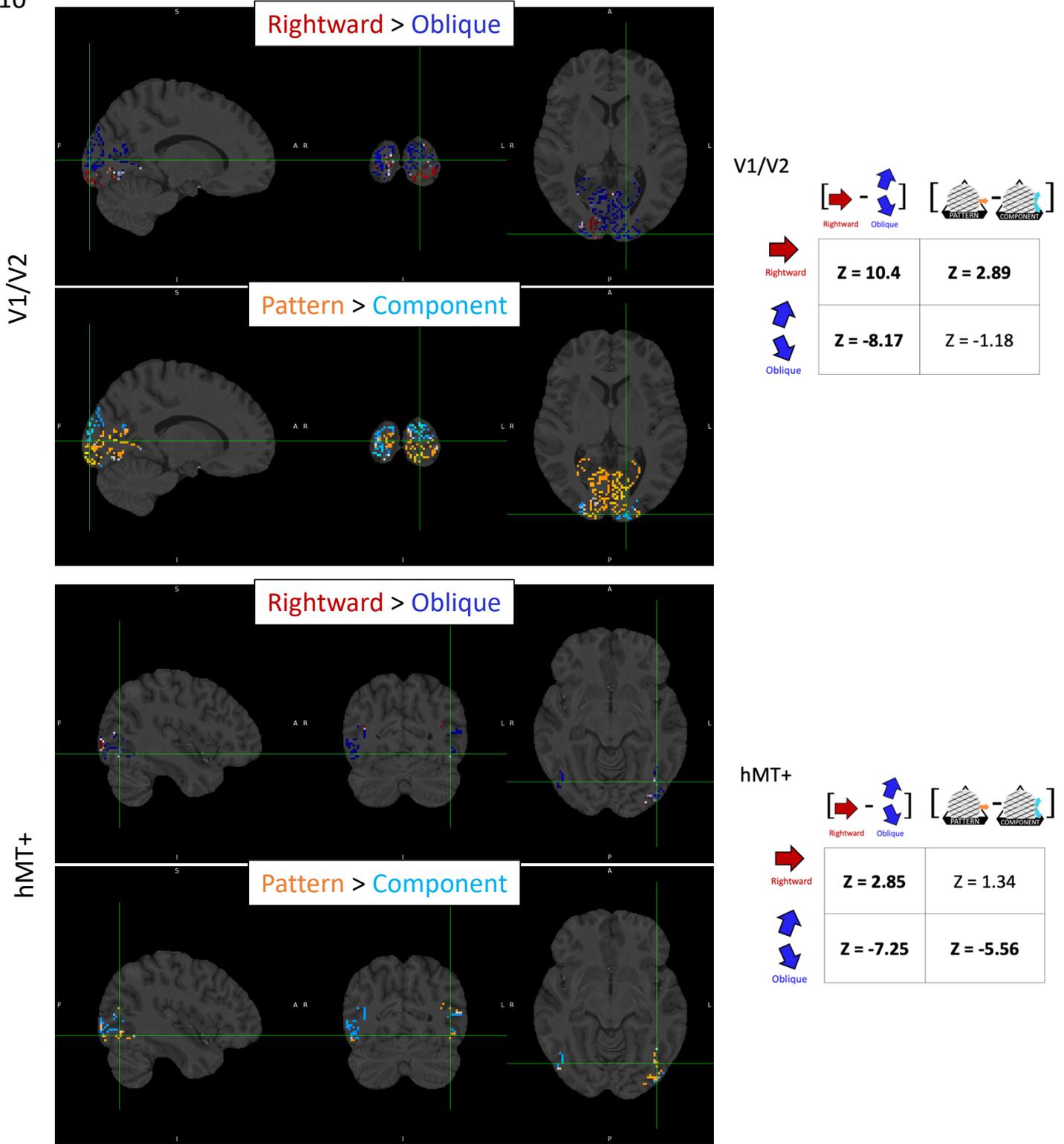


Figure S1.10: Direction-selectivity and Perceptual decision maps in subject 10
Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significant different from zero: $p < 0.01$).

S11

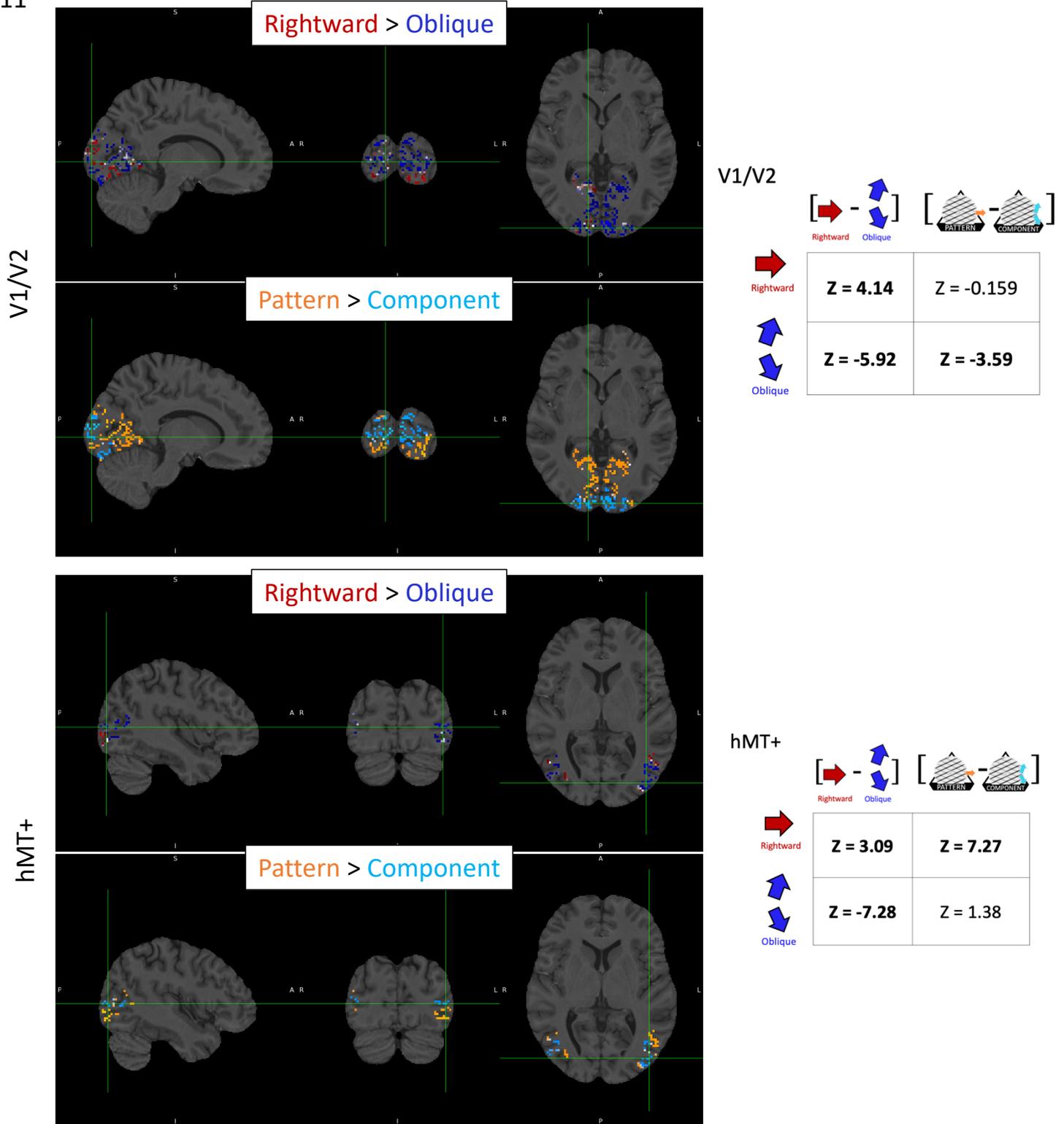


Figure S1.11: Direction-selectivity and Perceptual decision maps in subject 11
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S12

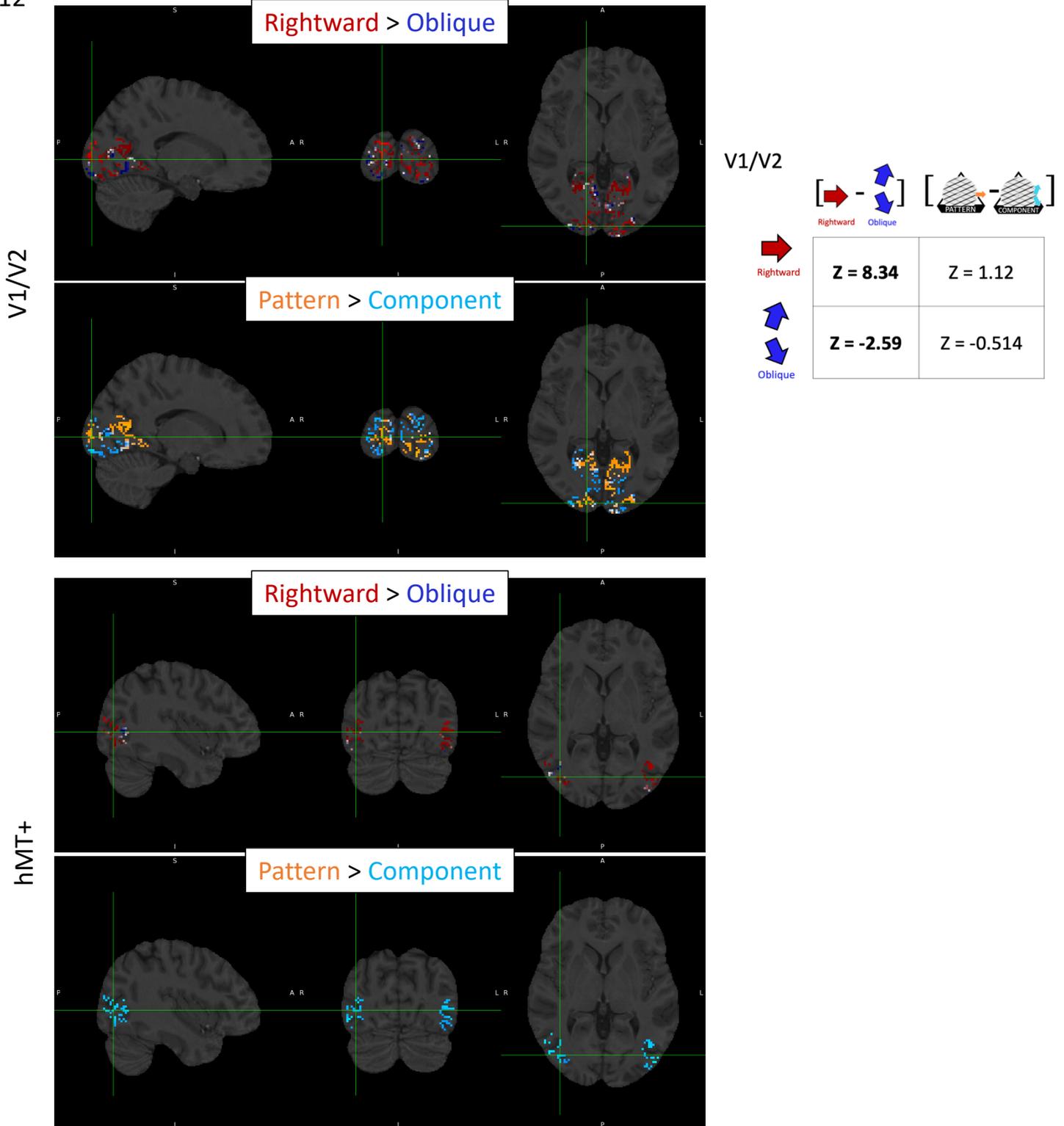


Figure S1.12 : Direction-selectivity and Perceptual decision maps in subject 12
 Localizer: In V1/V2, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 However, model failed in detecting Oblique selectivity in hMT+.
 Bistable motion: In V1/V2, Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 hMT+ predominantly responded to Component state.
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S13

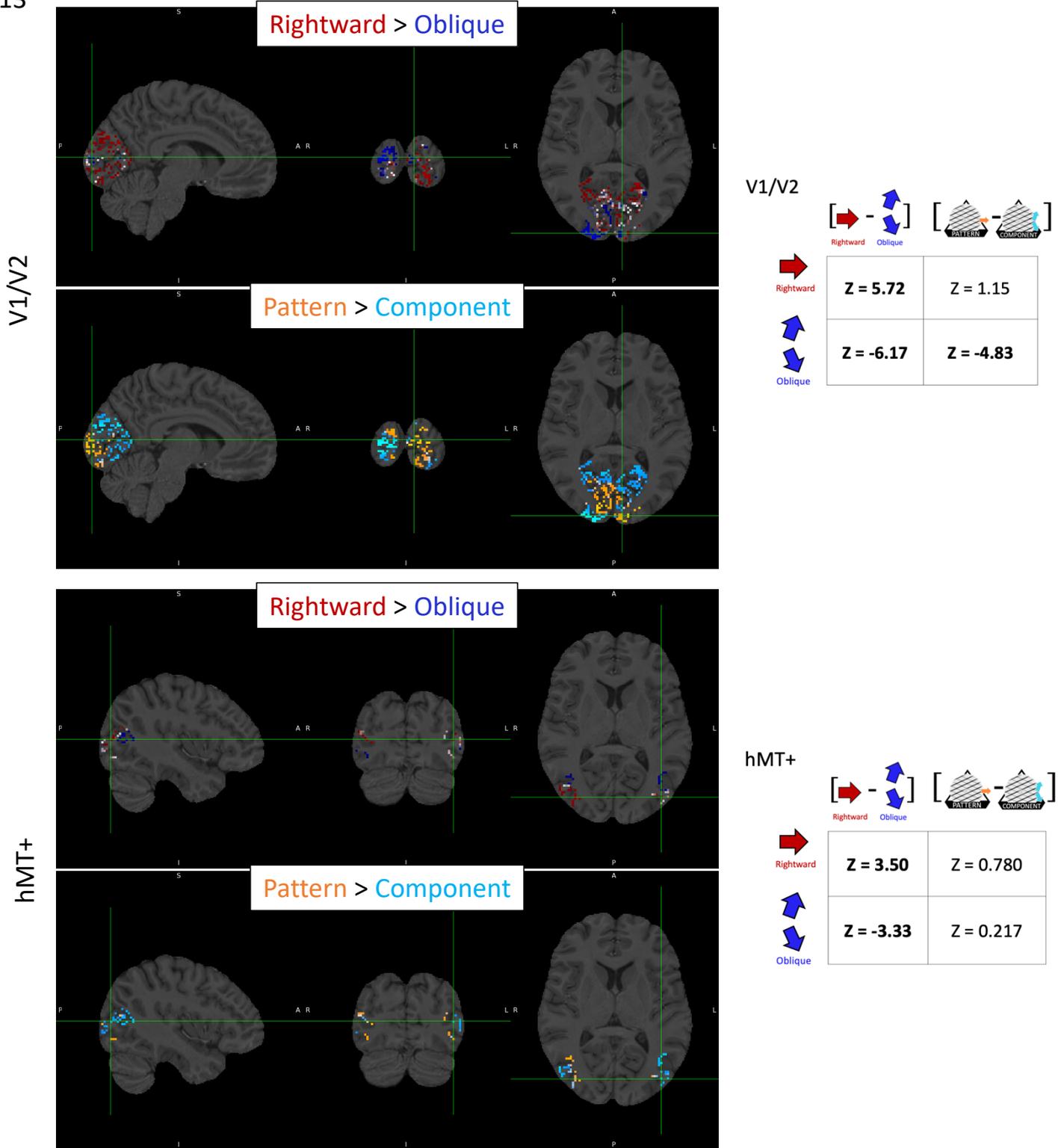


Figure S1.13: Direction-selectivity and Perceptual decision maps in subject 13
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S14

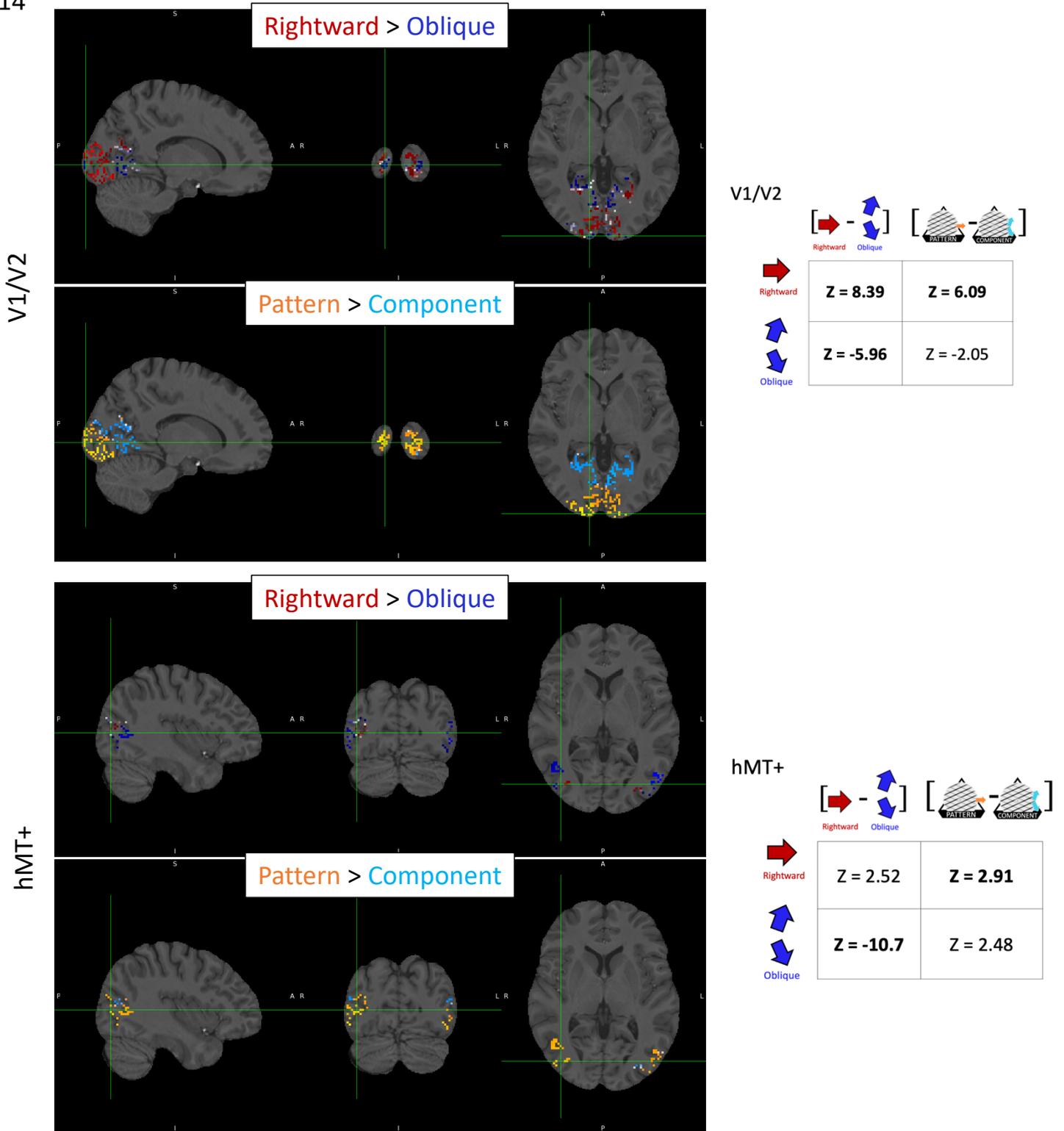


Figure S1.14: Direction-selectivity and Perceptual decision maps in subject 14
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S15

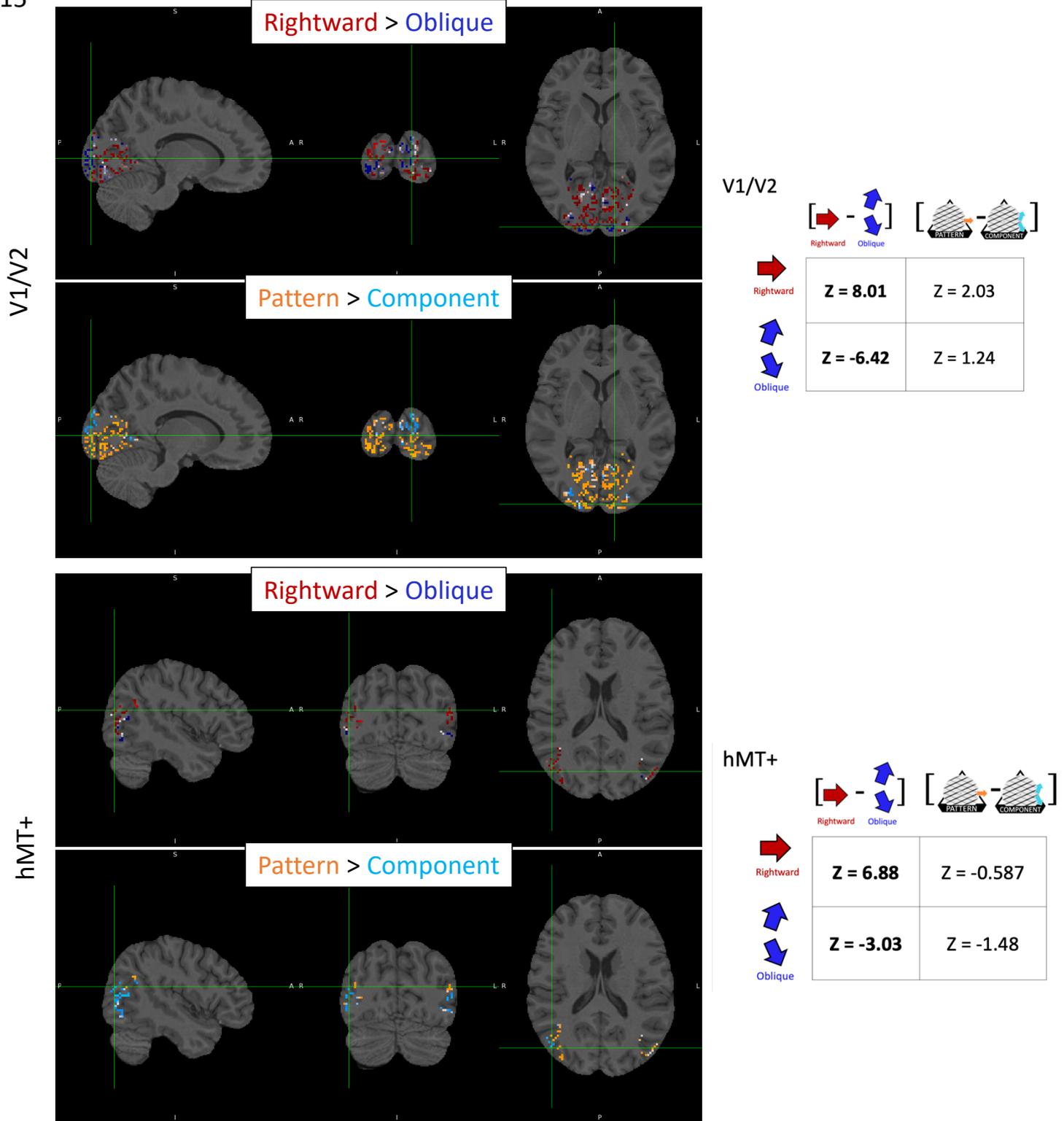


Figure S1.15: Direction-selectivity and Perceptual decision maps in subject 15

Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).

Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).

Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S16

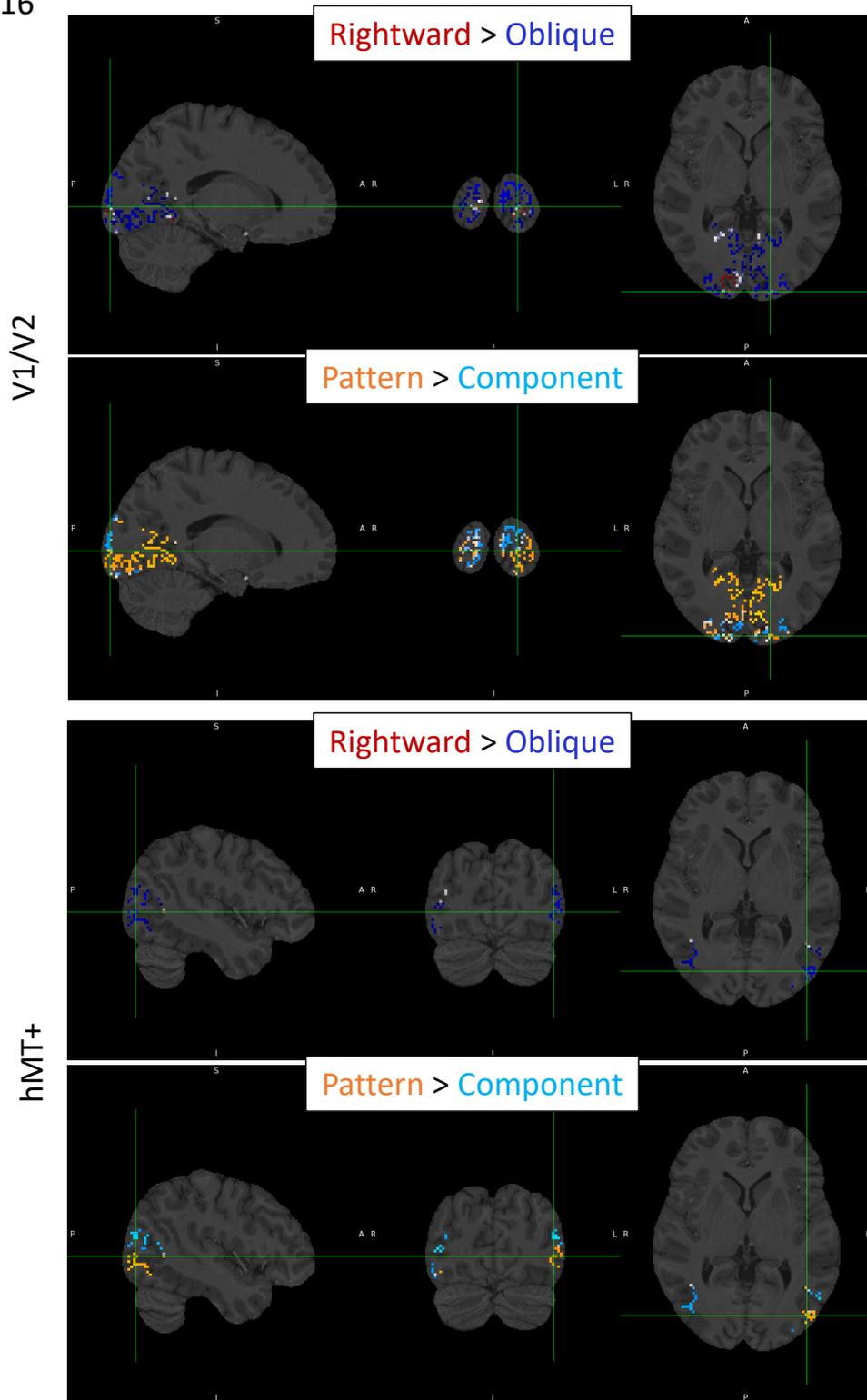


Figure S1.16: Direction-selectivity and Perceptual decision maps in subject 16
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed predominant response to Oblique directions (blue) ($p < 0.01$ uncorr) and model failed to detect Rightward selectivity.
 Bistable motion: Overall, Pattern > Component contrast resulted in both Component and Pattern-selective (in cyan) responses ($p < 0.01$ uncorr).
 This subject was discarded from the group-level analyses because of the undetected subdomains.

S17

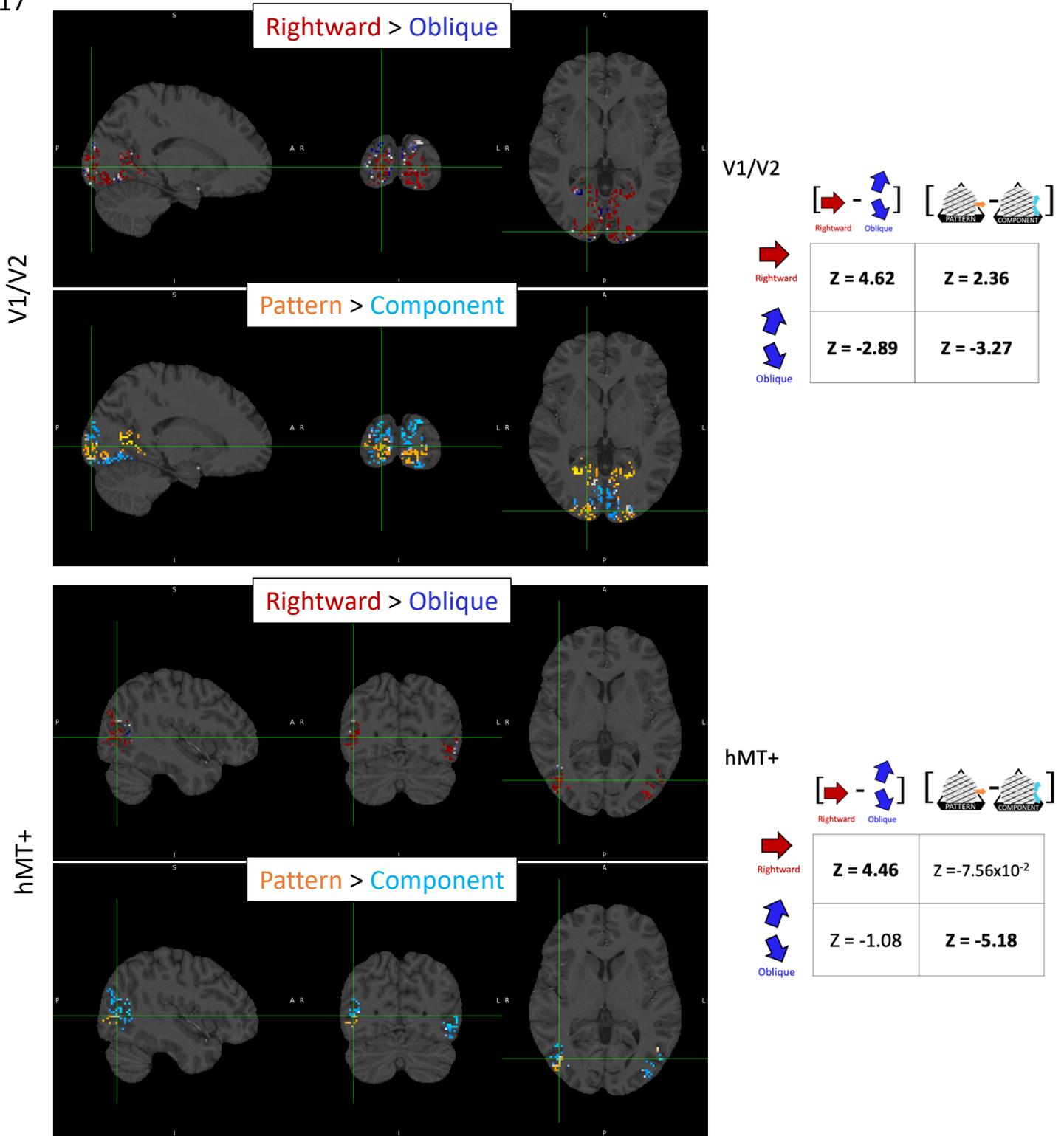
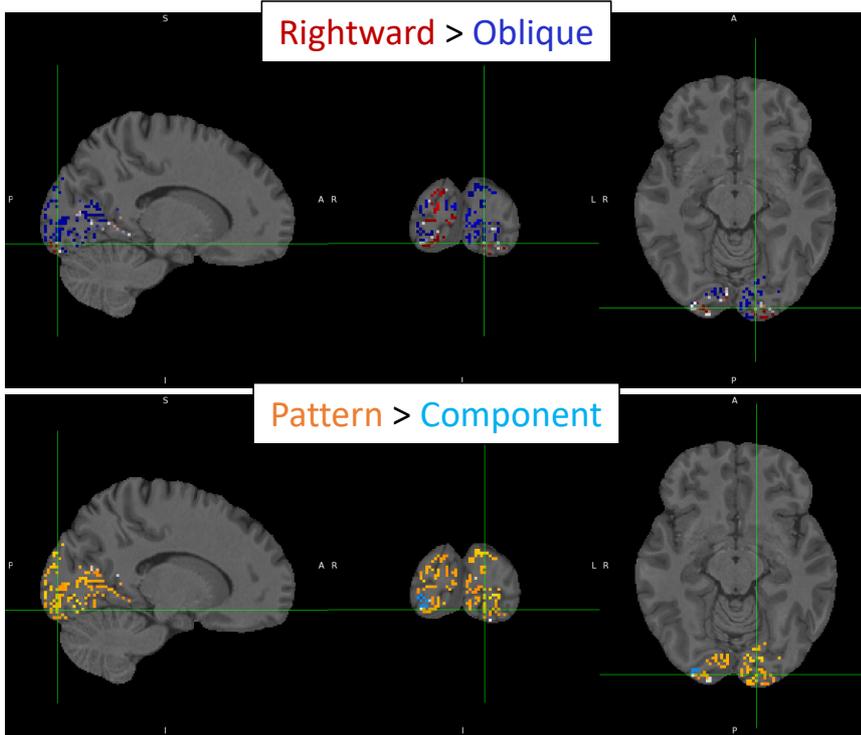
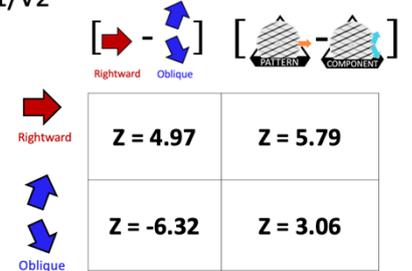


Figure S1.17: Direction-selectivity and Perceptual decision maps in subject 17
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

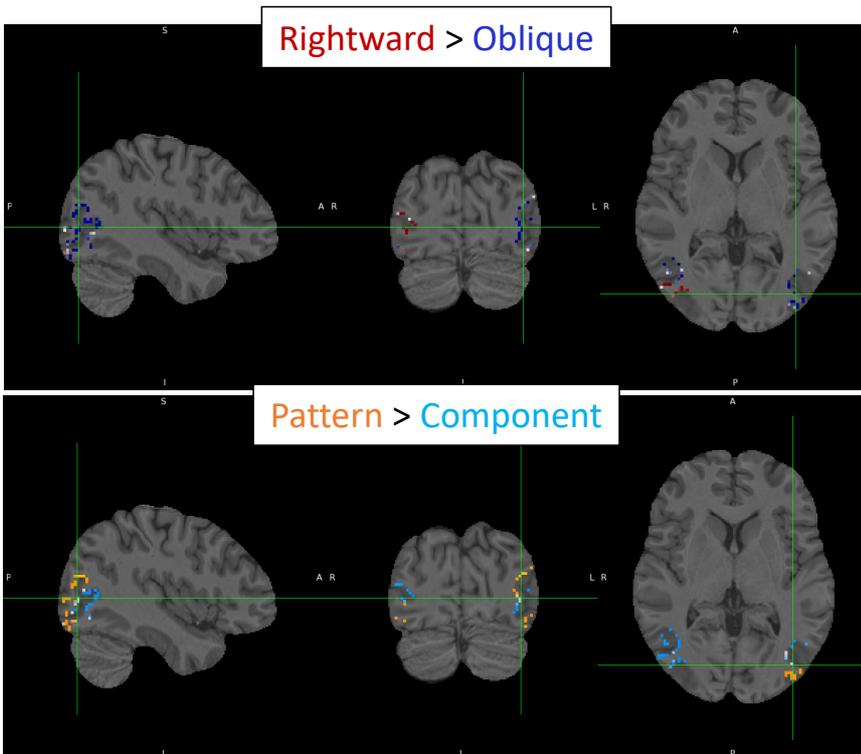
V1/V2



V1/V2



hMT+



hMT+

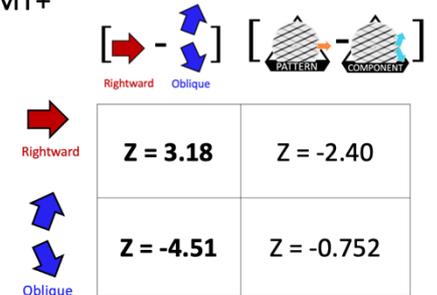
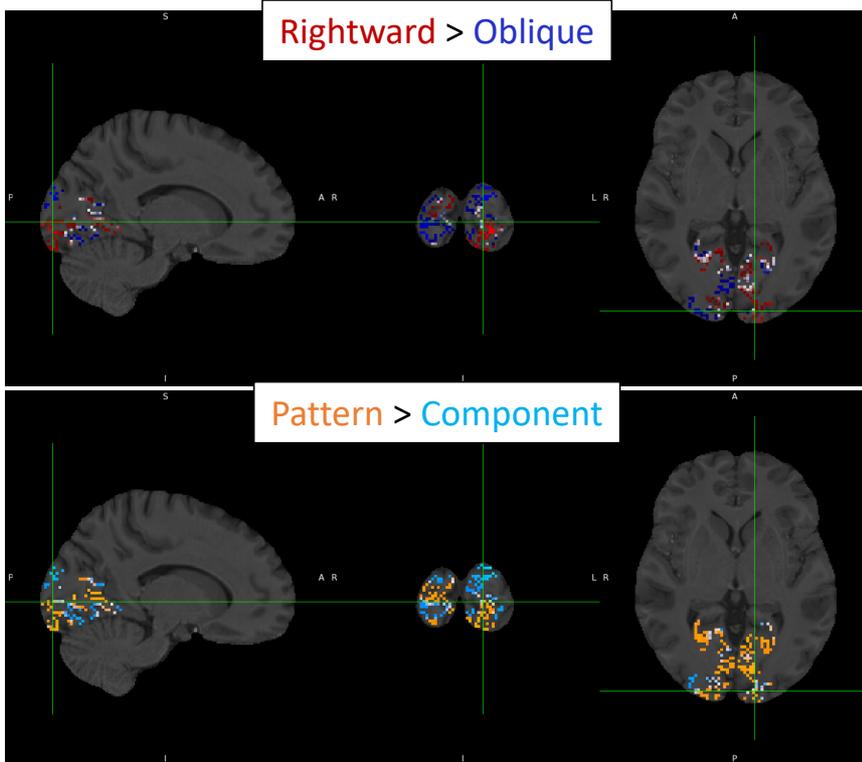


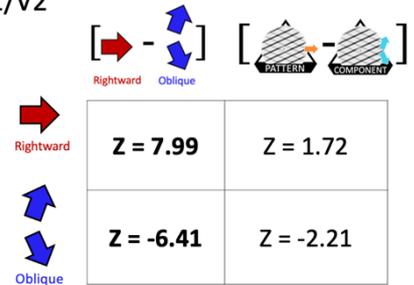
Figure S1.18: Direction-selectivity and Perceptual decision maps in subject 18
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S19

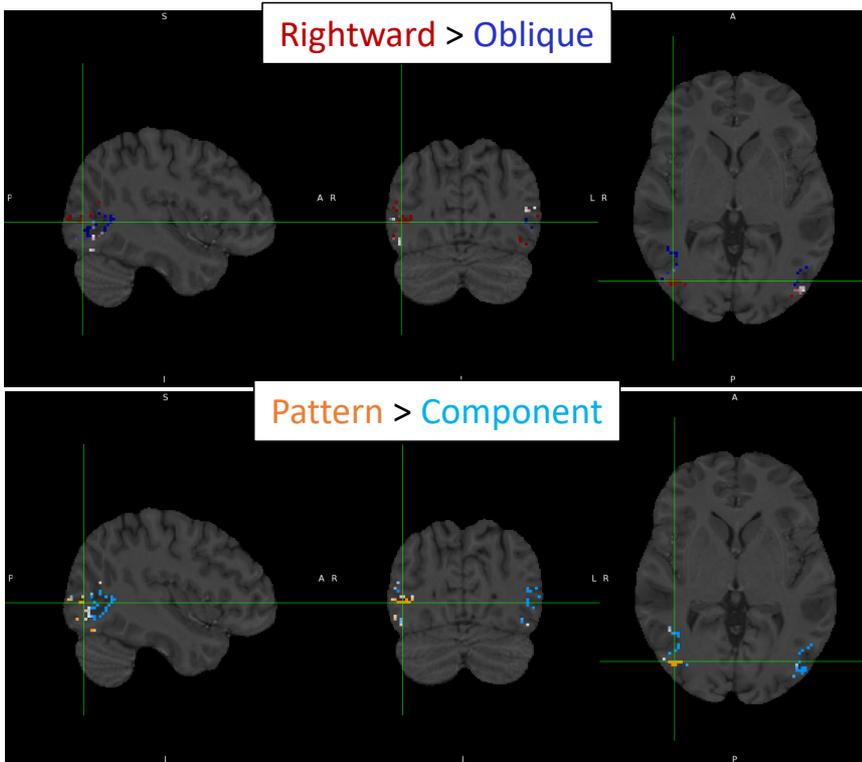
V1/V2



V1/V2



hMT+



hMT+

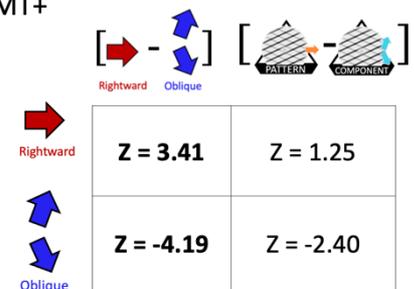
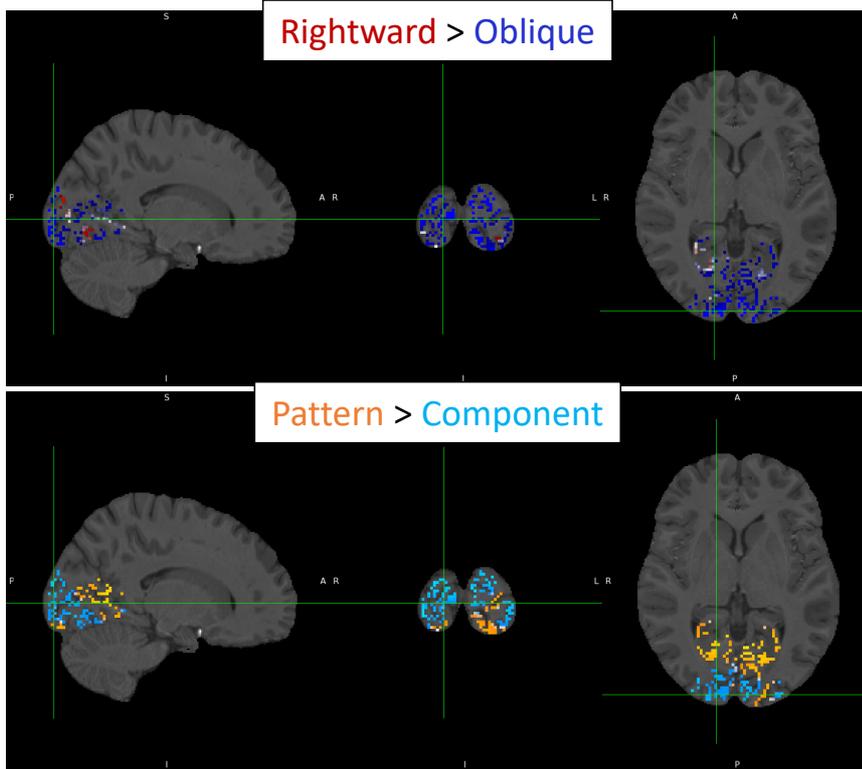


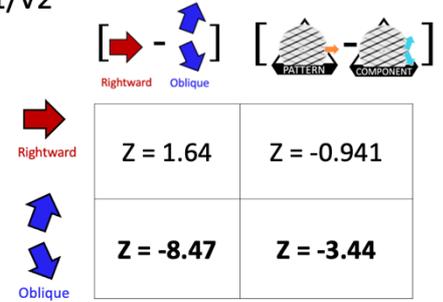
Figure S1.19: Direction-selectivity and Perceptual decision maps in subject 19
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S20

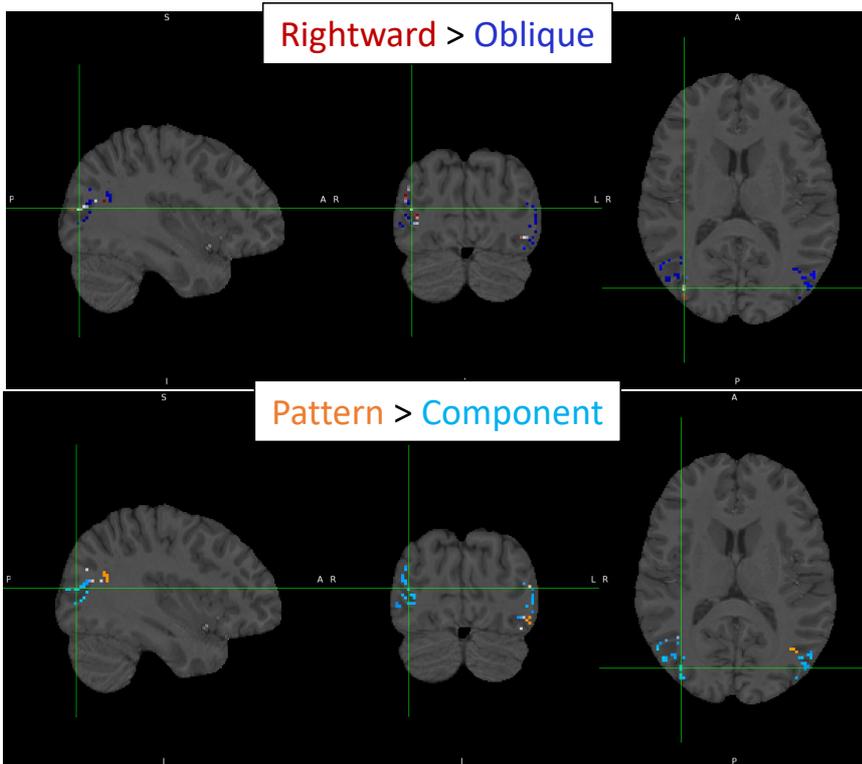
V1/V2



V1/V2



hMT+



hMT+

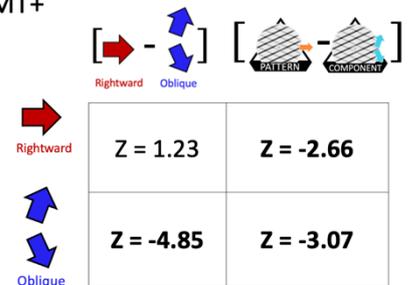
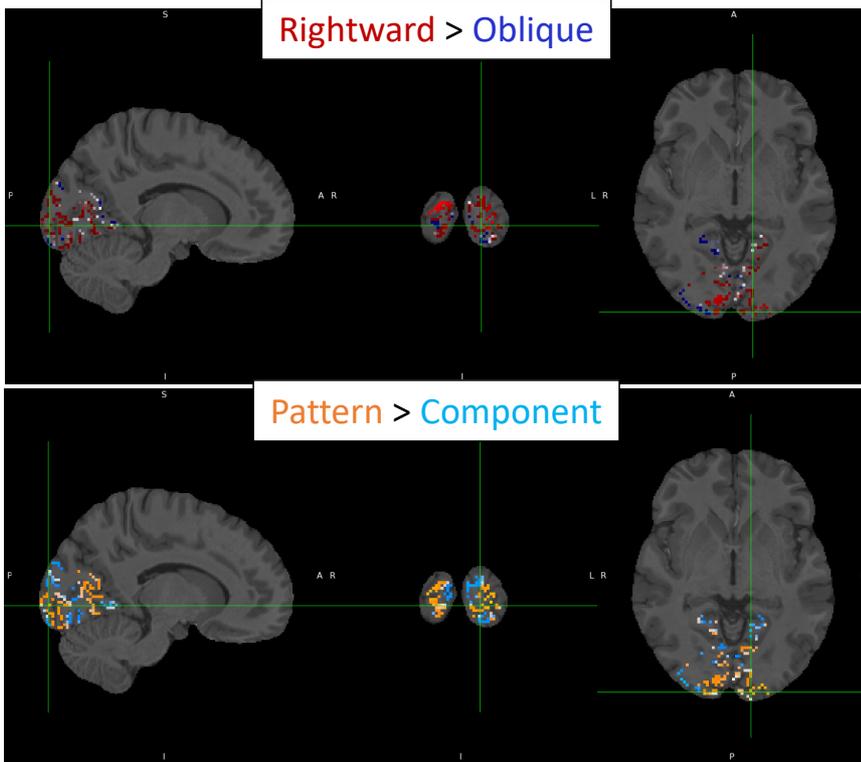


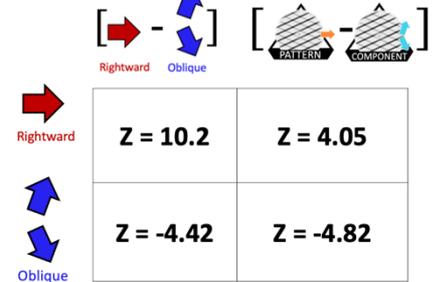
Figure S1.20: Direction-selectivity and Perceptual decision maps in subject 20
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S21

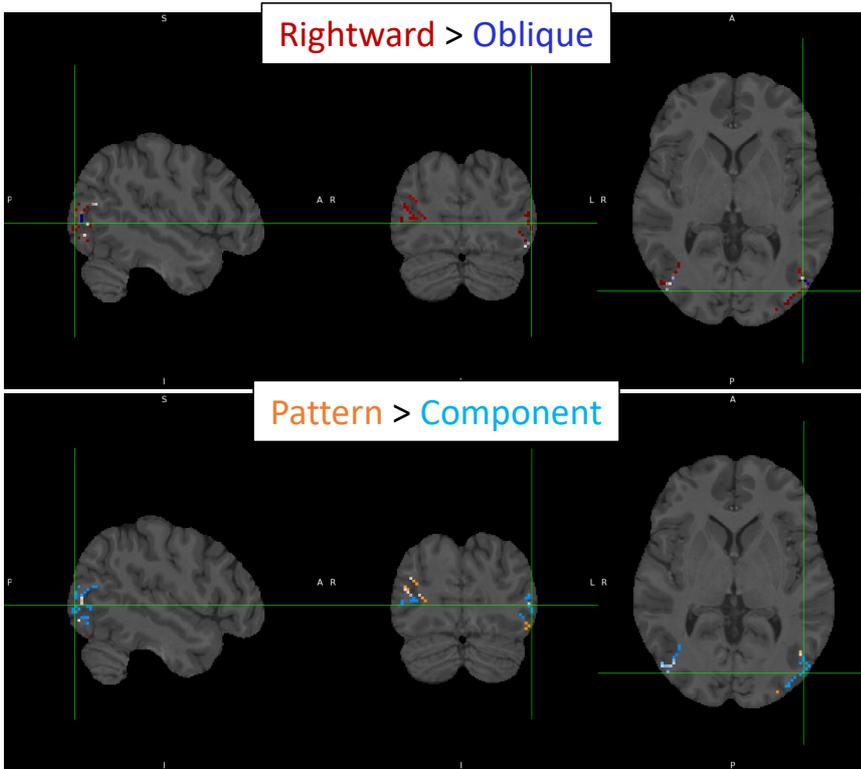
V1/V2



V1/V2



hMT+



hMT+

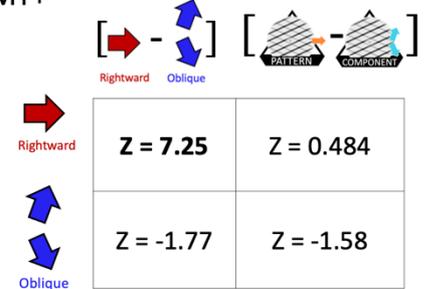
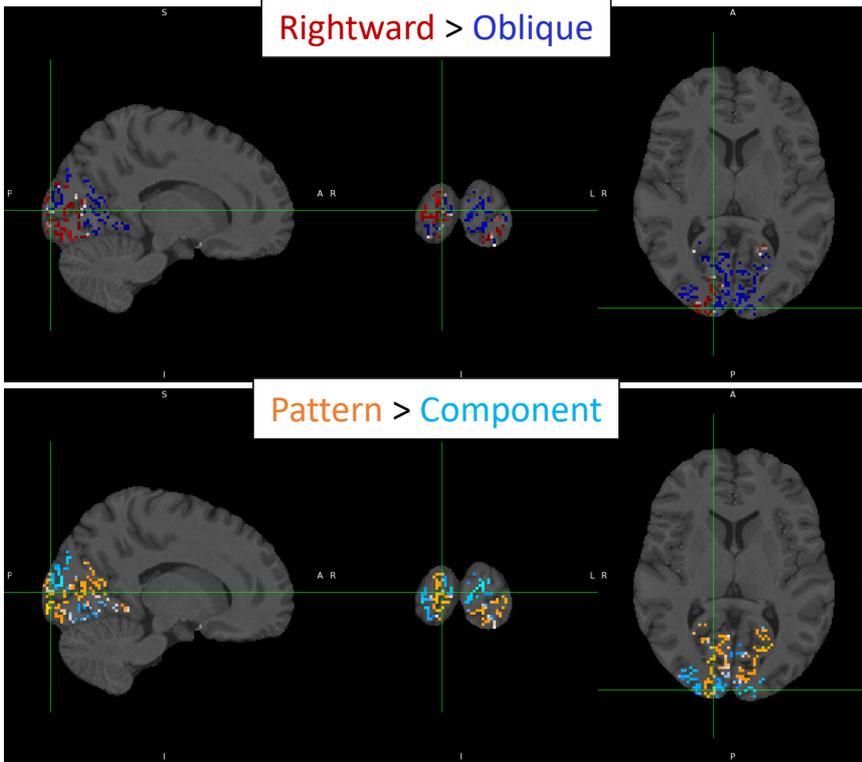


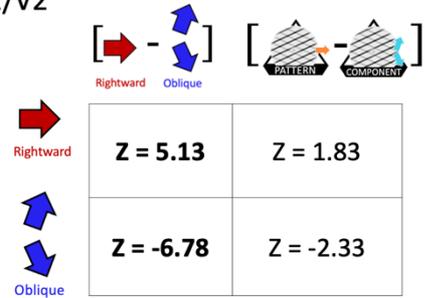
Figure SI.21: Direction-selectivity and Perceptual decision maps in subject 21
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S22

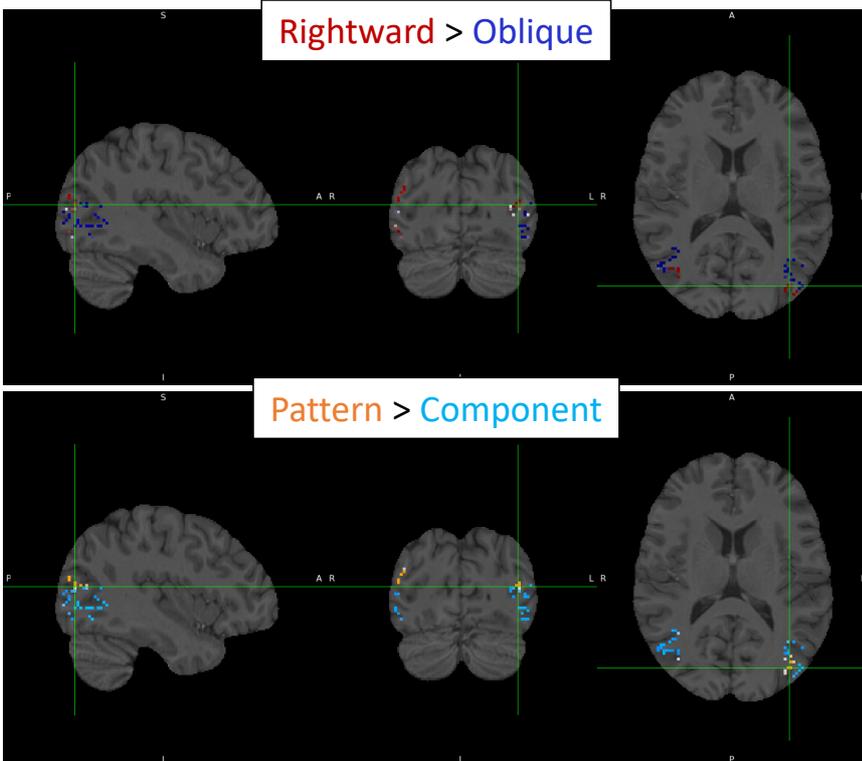
V1/V2



V1/V2



hMT+



hMT+

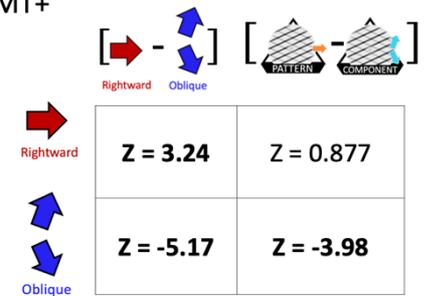
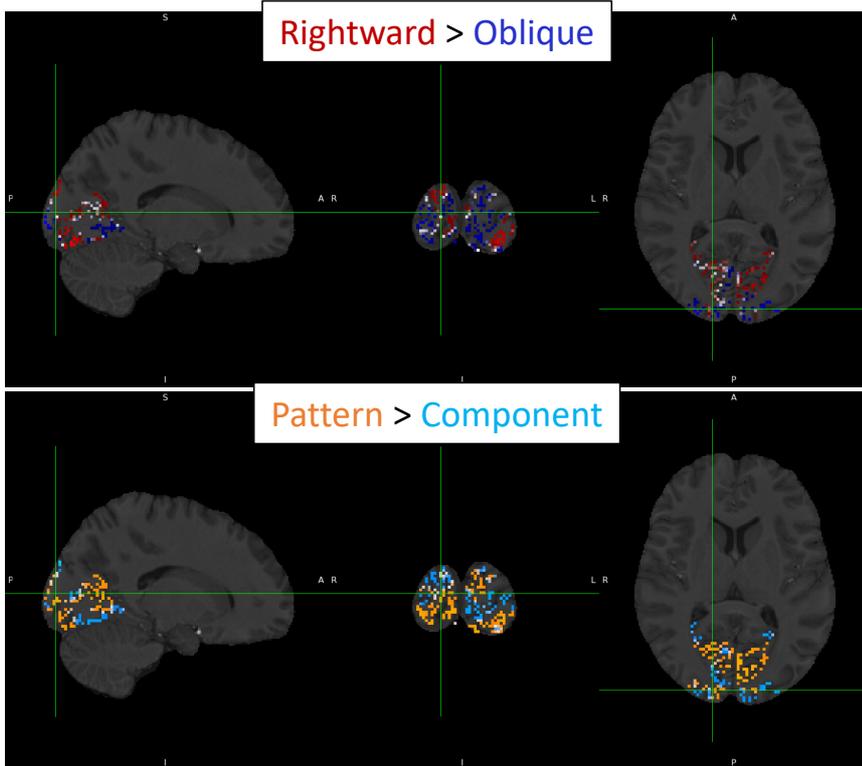


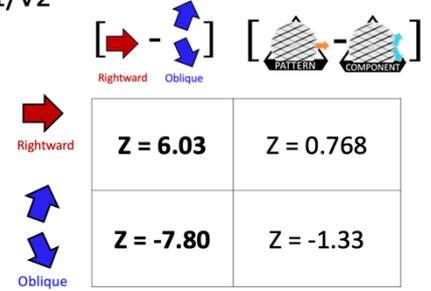
Figure S1.22: Direction-selectivity and Perceptual decision maps in subject 22
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S23

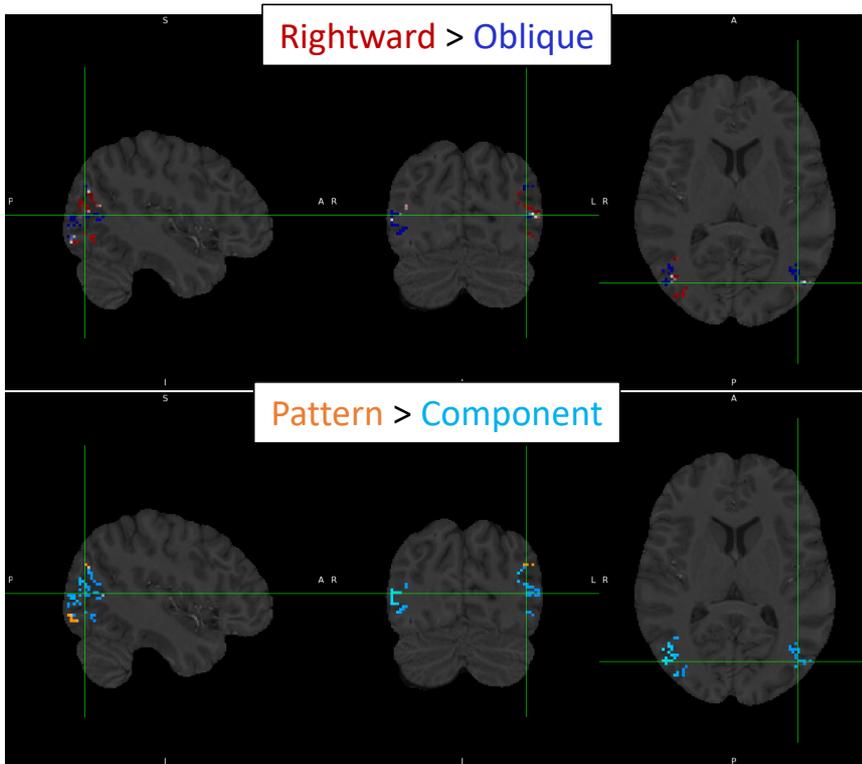
V1/V2



V1/V2



hMT+



hMT+

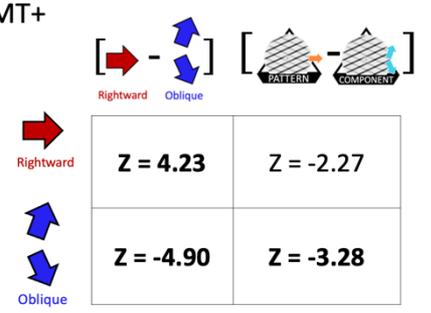


Figure S1.23: Direction-selectivity and Perceptual decision maps in subject 23

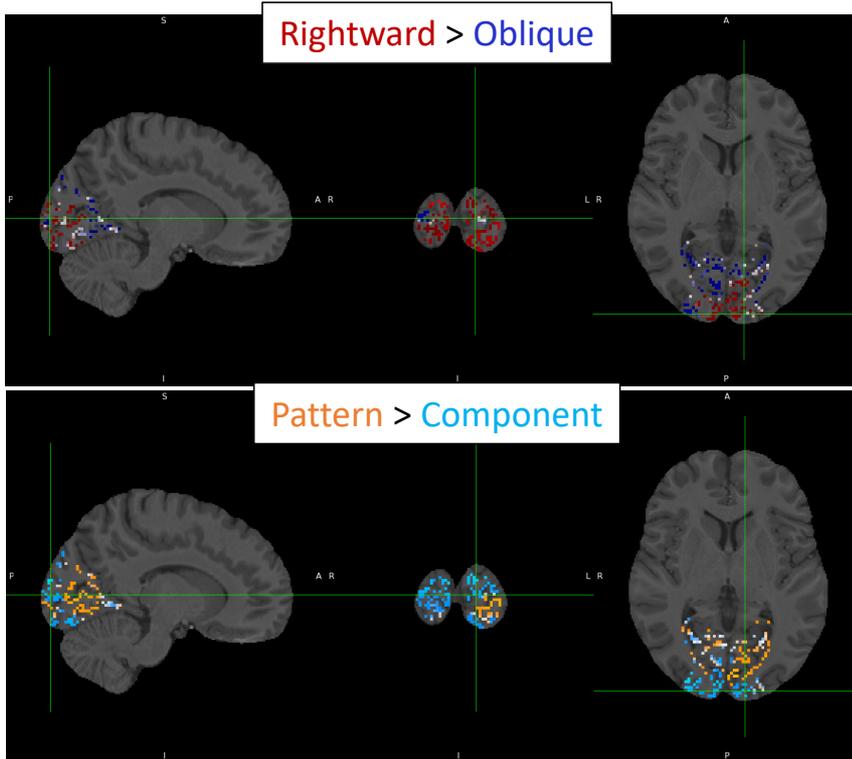
Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).

Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).

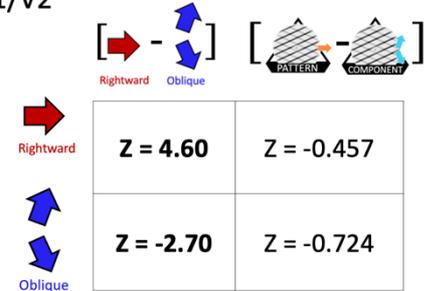
Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S24

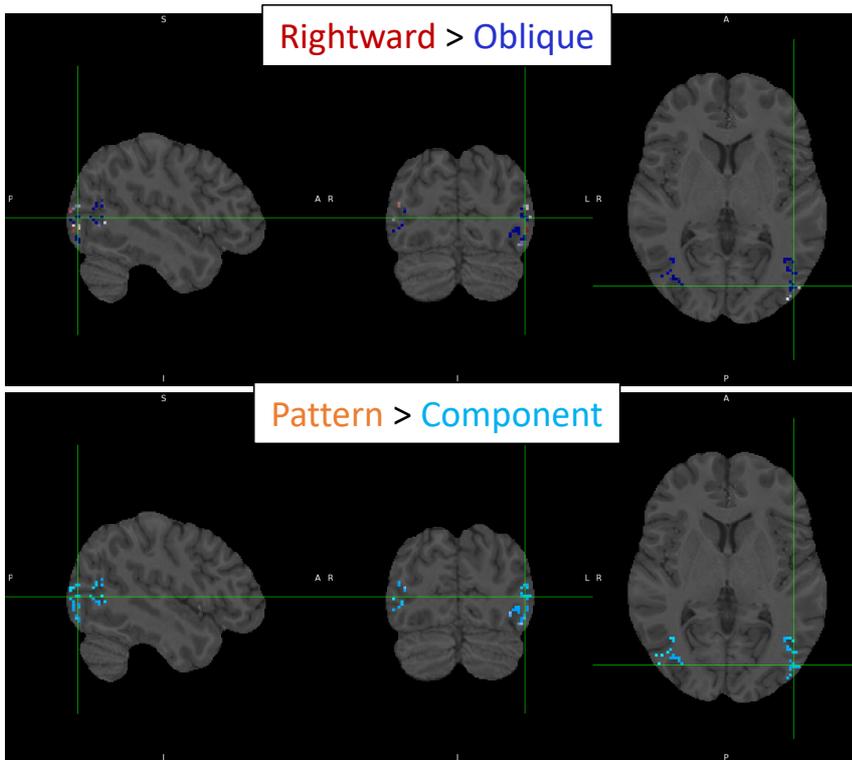
V1/V2



V1/V2



hMT+



hMT+

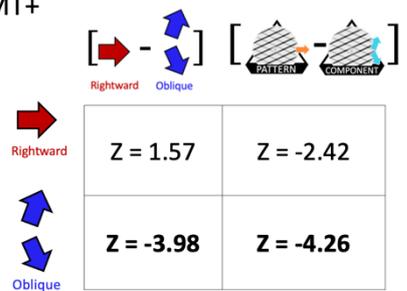
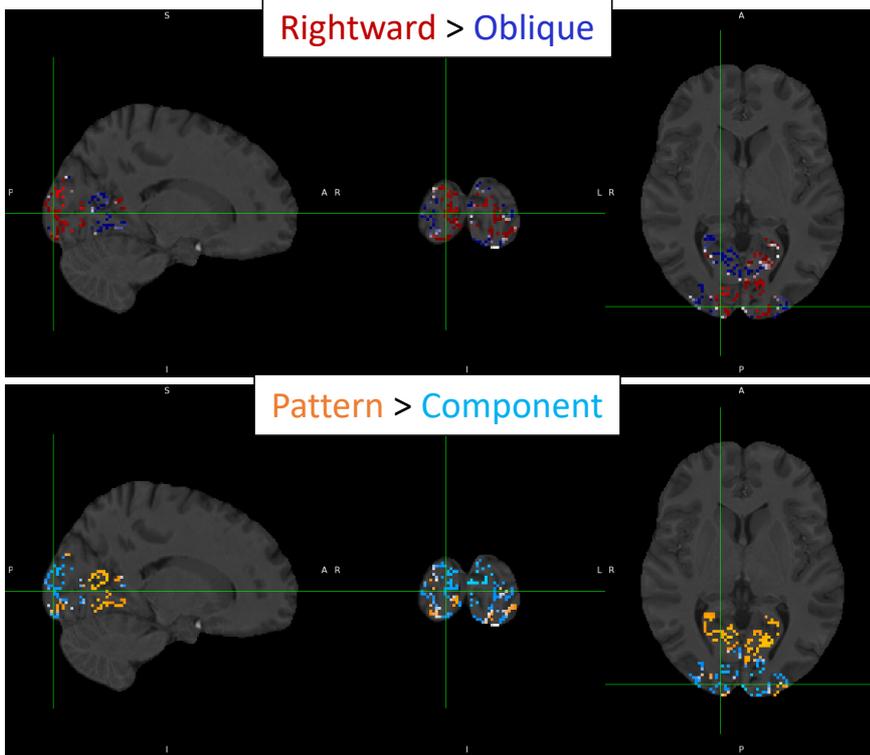


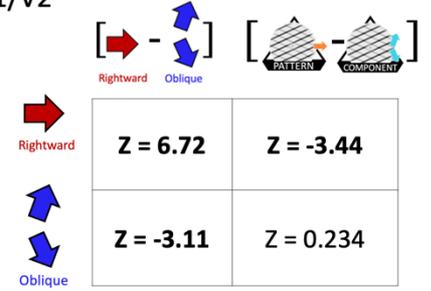
Figure S1.24: Direction-selectivity and Perceptual decision maps in subject 24
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S25

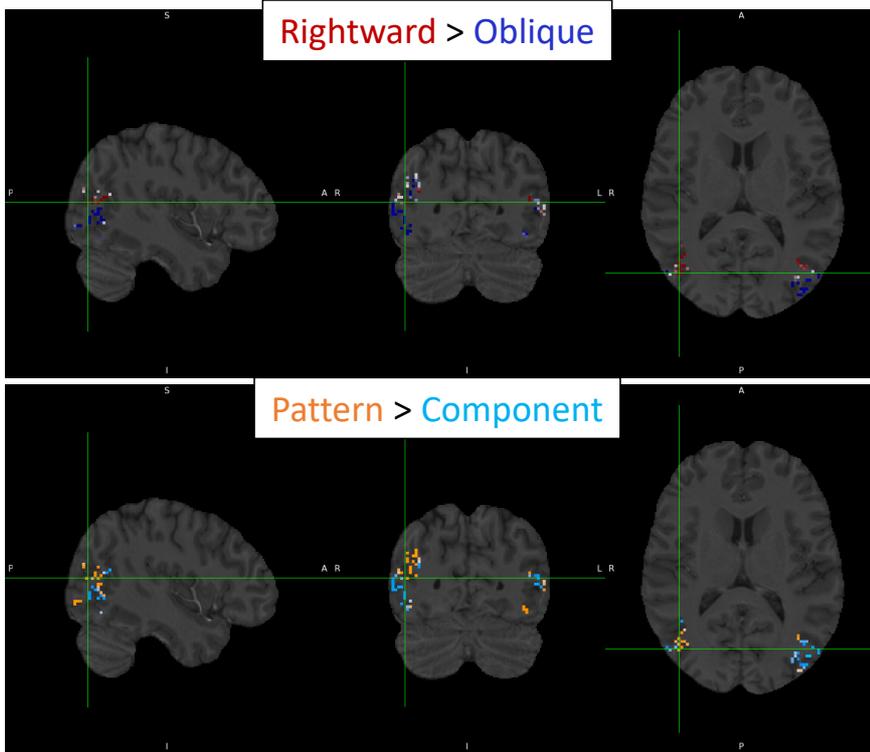
V1/V2



V1/V2



hMT+



hMT+

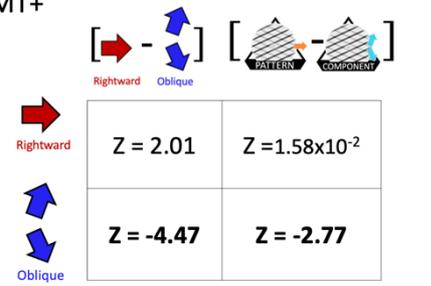
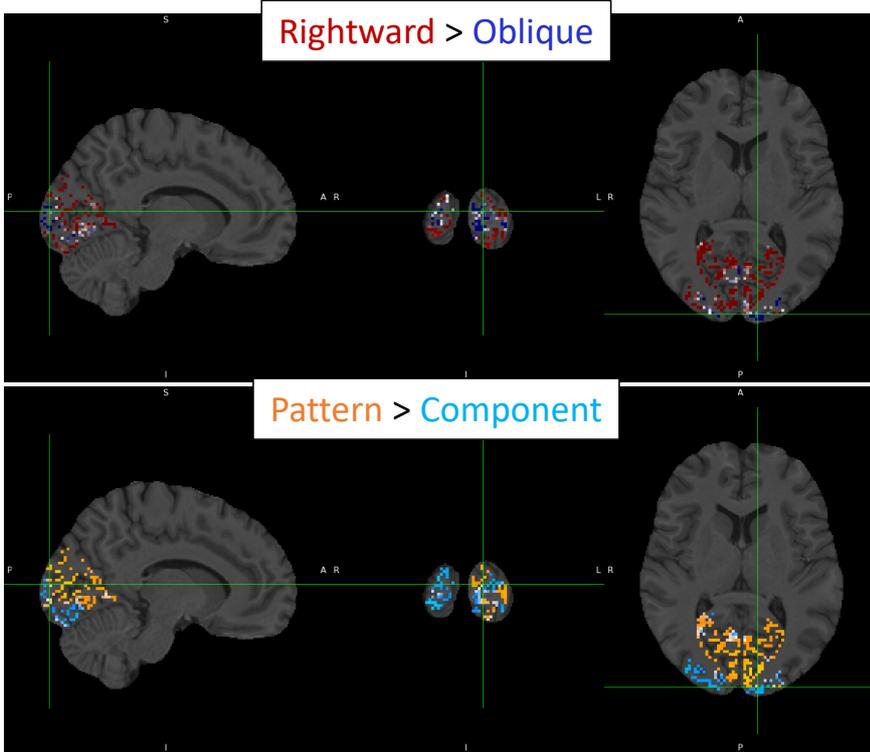


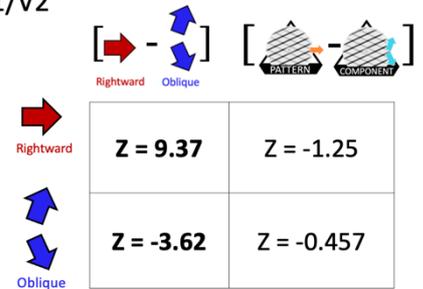
Figure S1.25: Direction-selectivity and Perceptual decision maps in subject 25
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S26

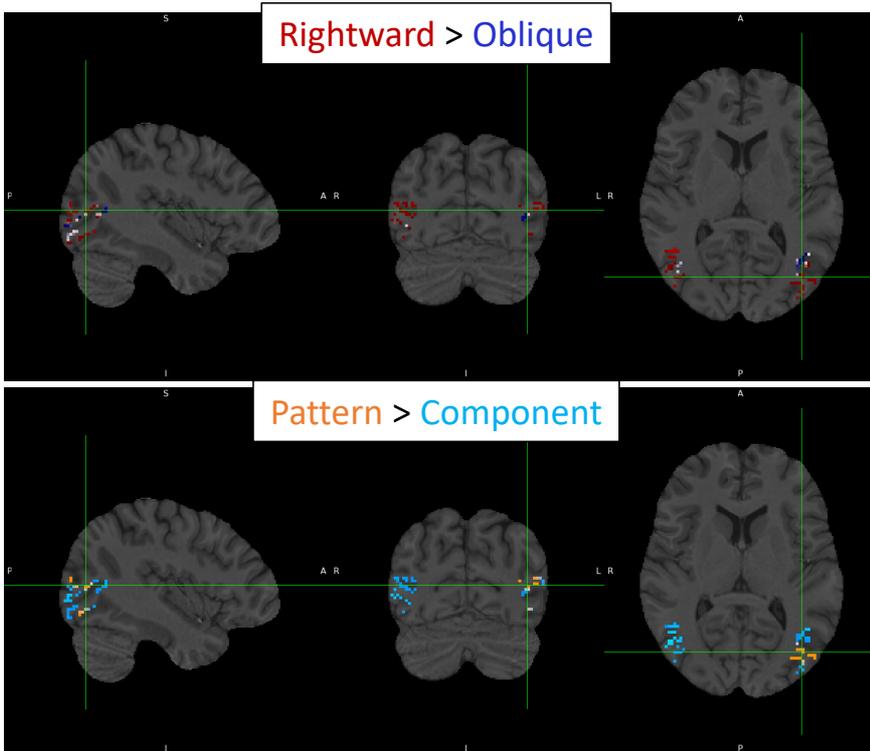
V1/V2



V1/V2



hMT+



hMT+

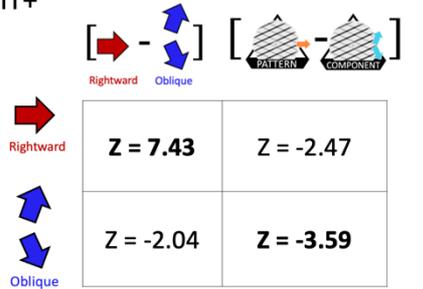


Figure S1.26: Direction-selectivity and Perceptual decision maps in subject 26
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

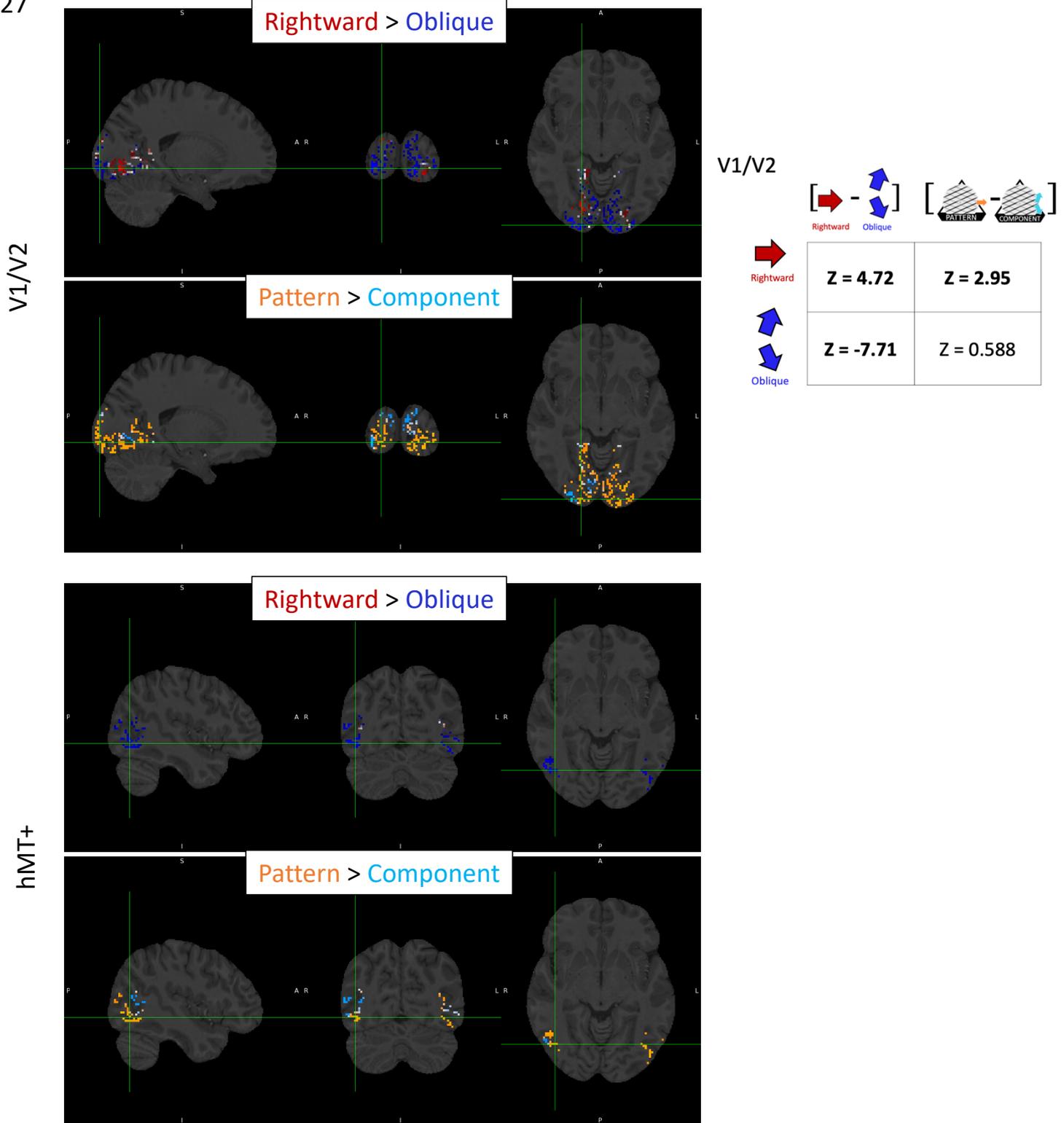
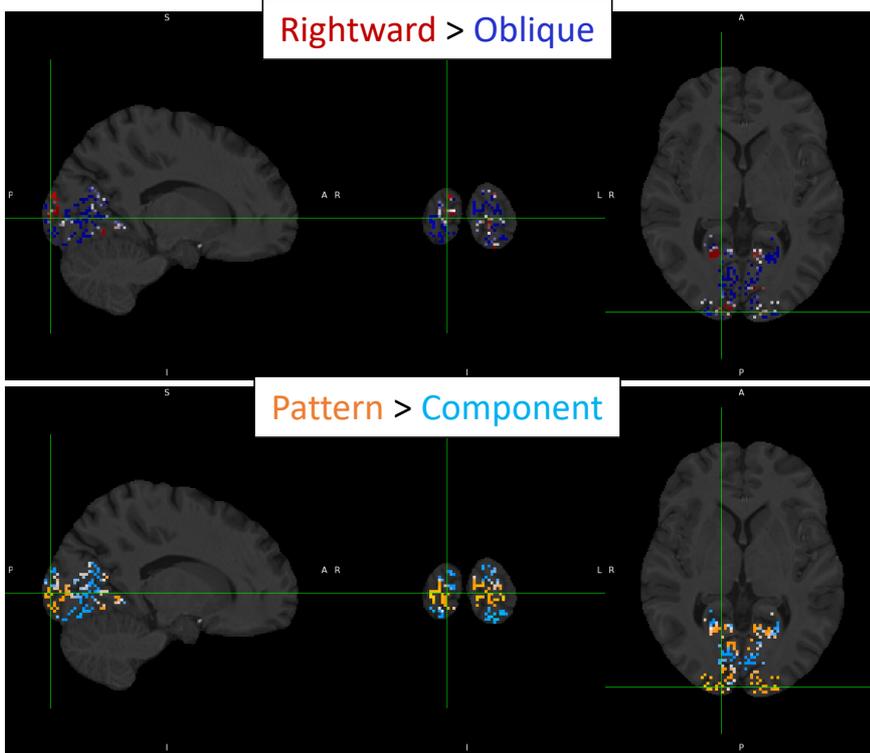
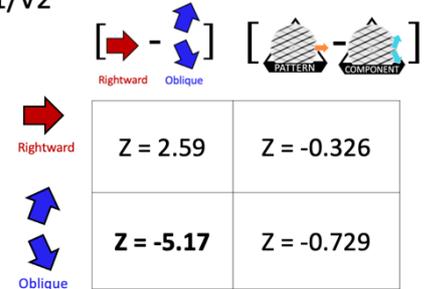


Figure S1.27: Direction-selectivity and Perceptual decision maps in subject 27
 Localizer: In V1/V2, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 However, model failed in detecting Rightward selectivity in hMT+.
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

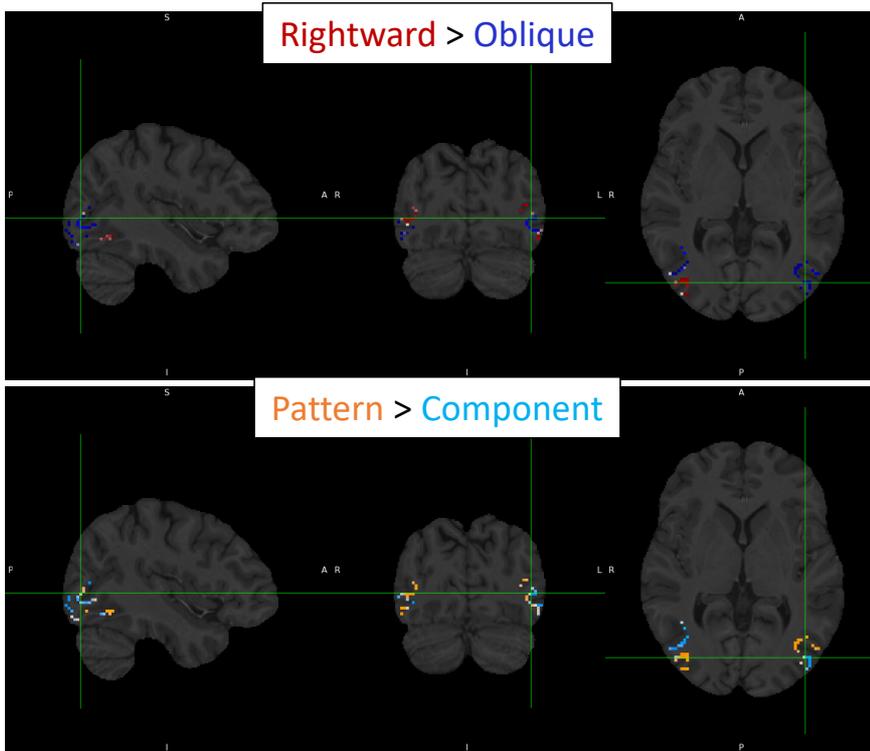
V1/V2



V1/V2



hMT+



hMT+

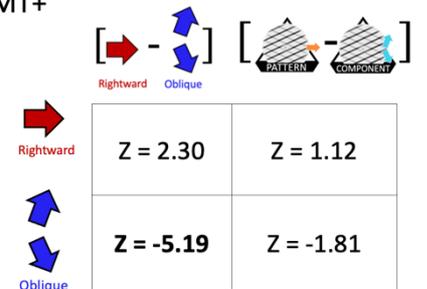
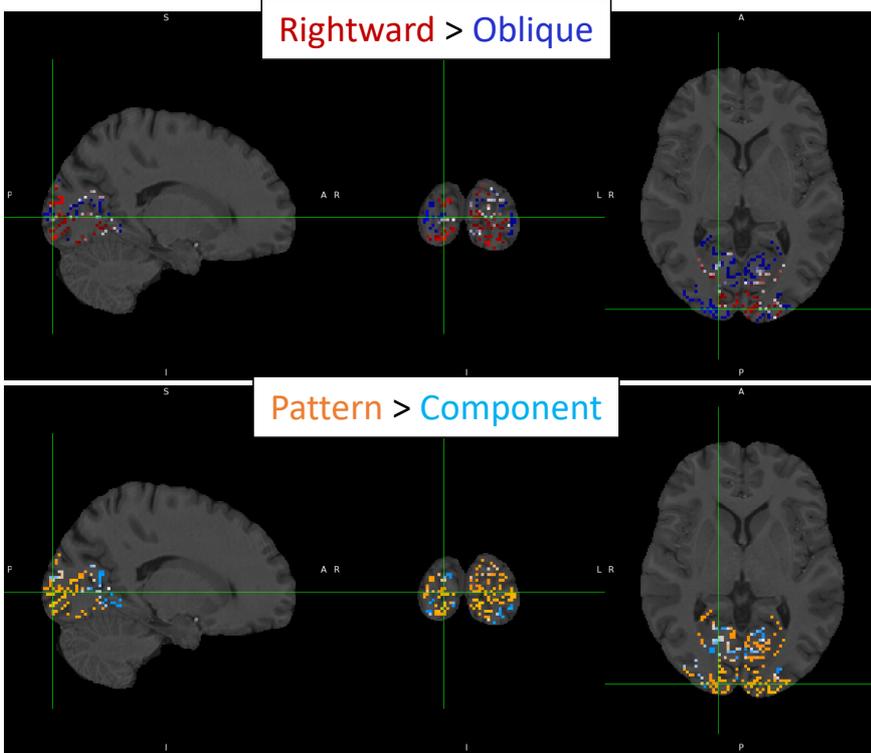


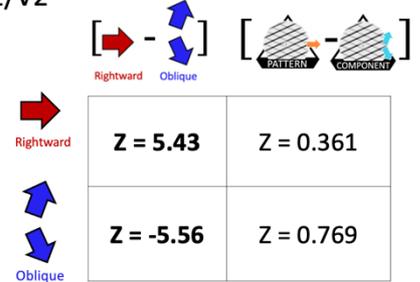
Figure S1.28: Direction-selectivity and Perceptual decision maps in subject 28
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S29

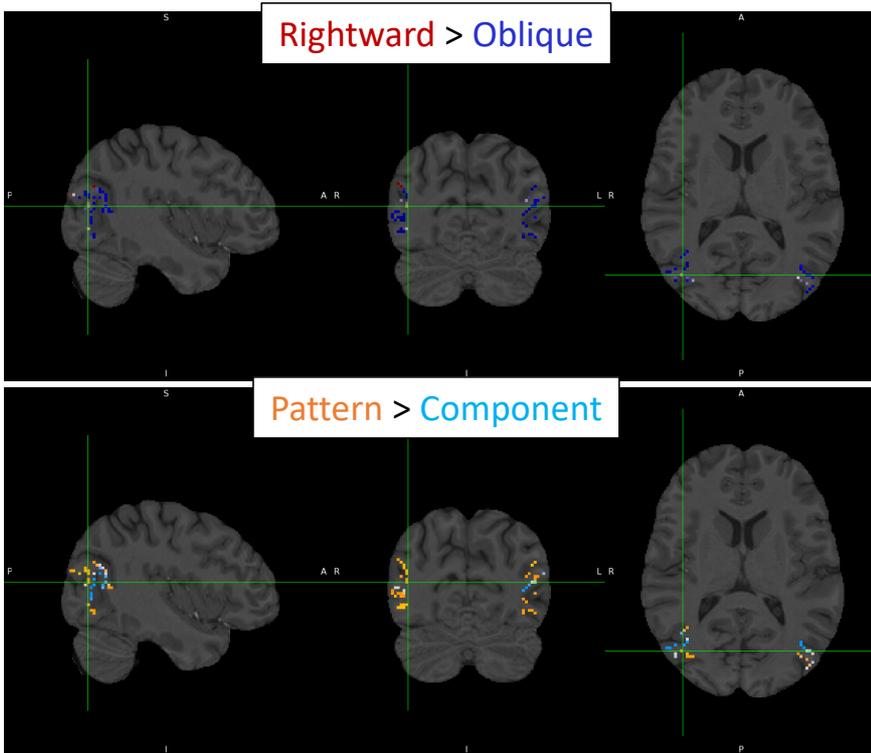
V1/V2



V1/V2



hMT+



hMT+

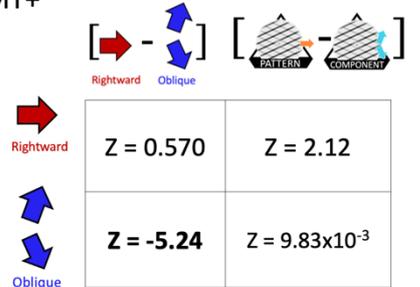
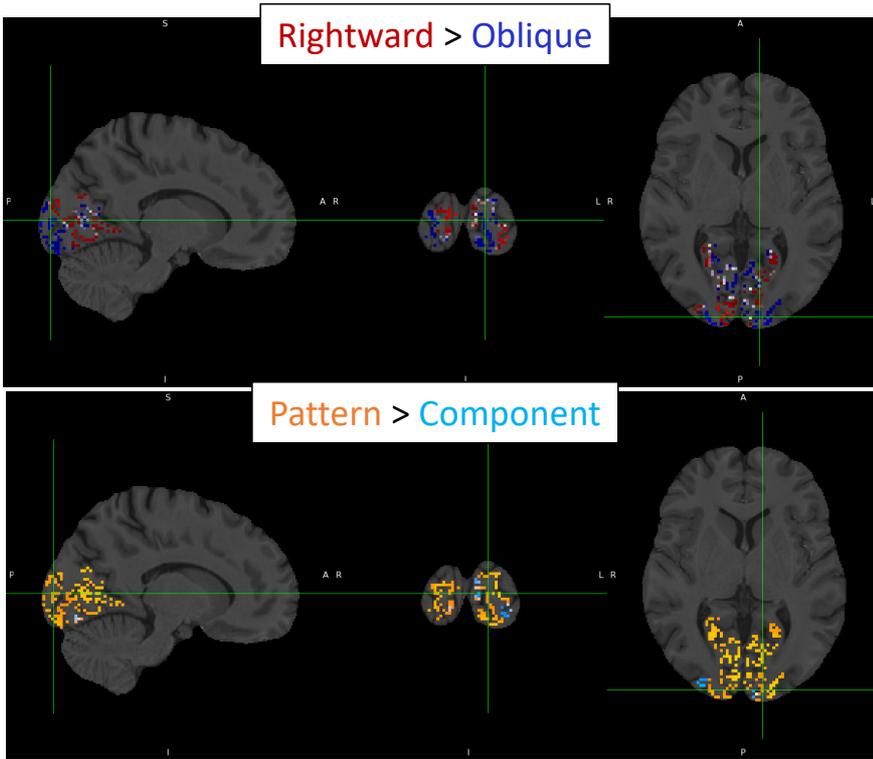


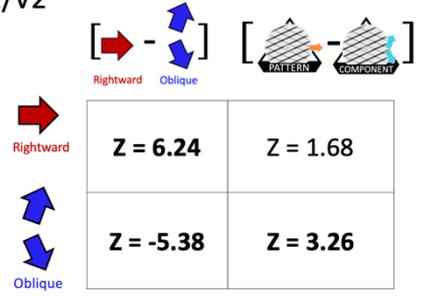
Figure S1.29: Direction-selectivity and Perceptual decision maps in subject 29
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S30

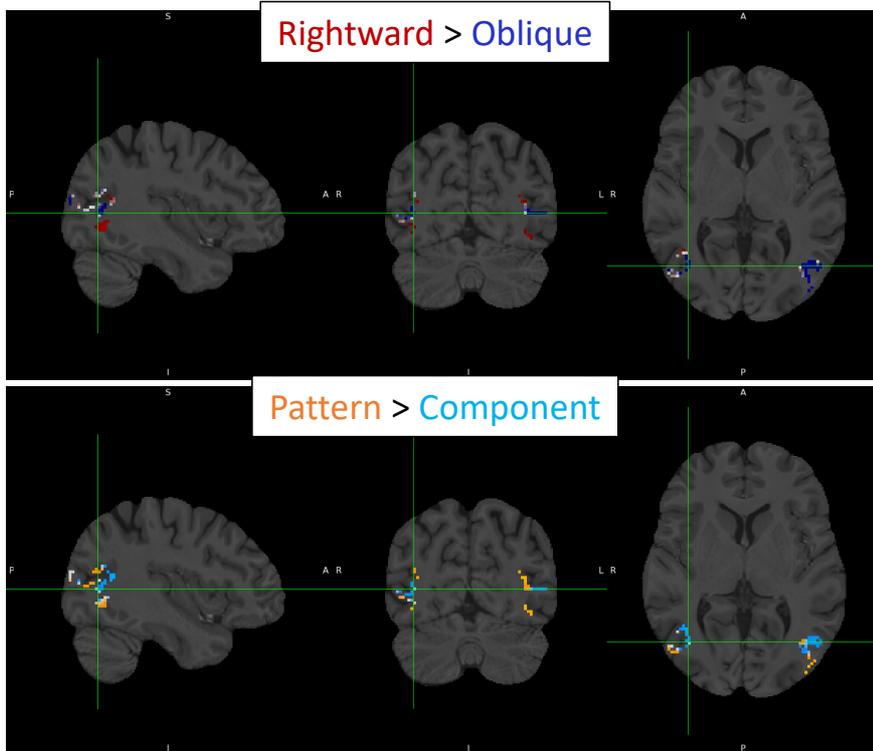
V1/V2



V1/V2



hMT+



hMT+

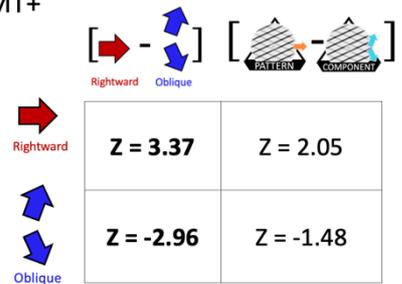
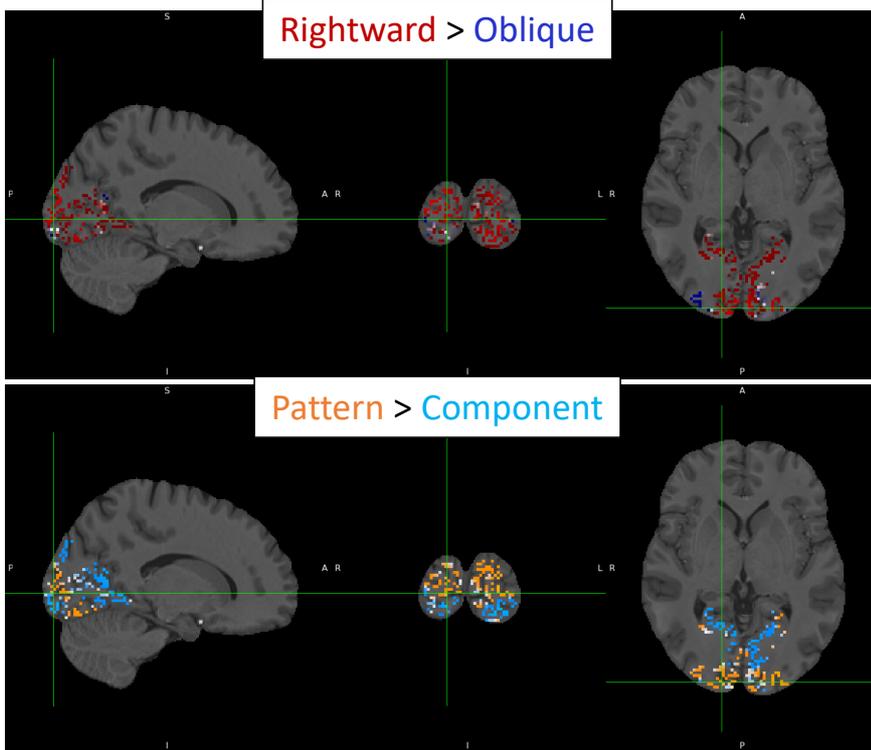


Figure S1.30: Direction-selectivity and Perceptual decision maps in subject 30
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

S31

V1/V2



hMT+

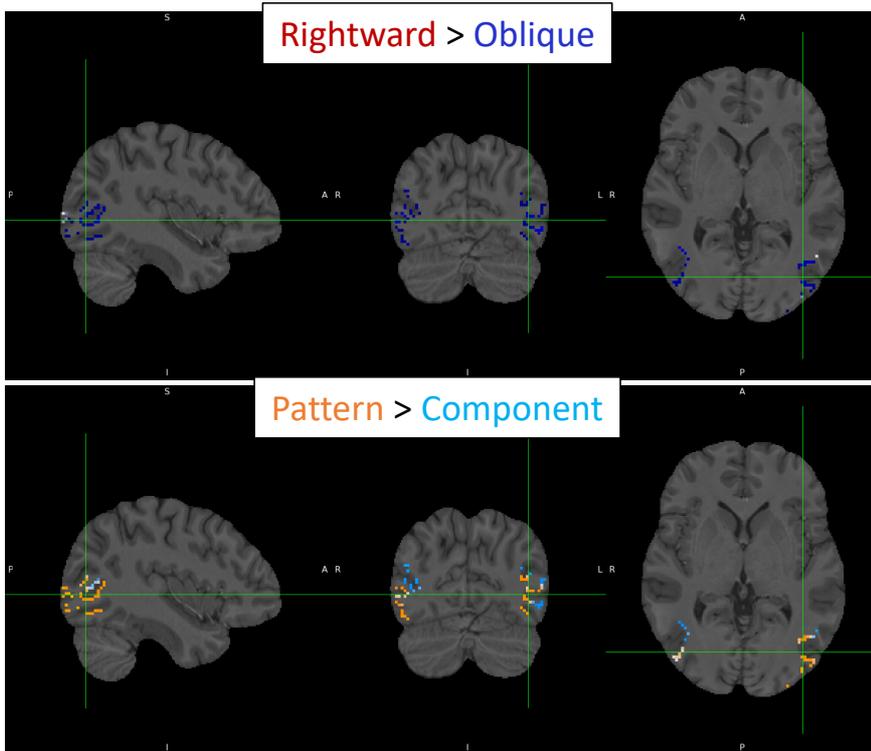
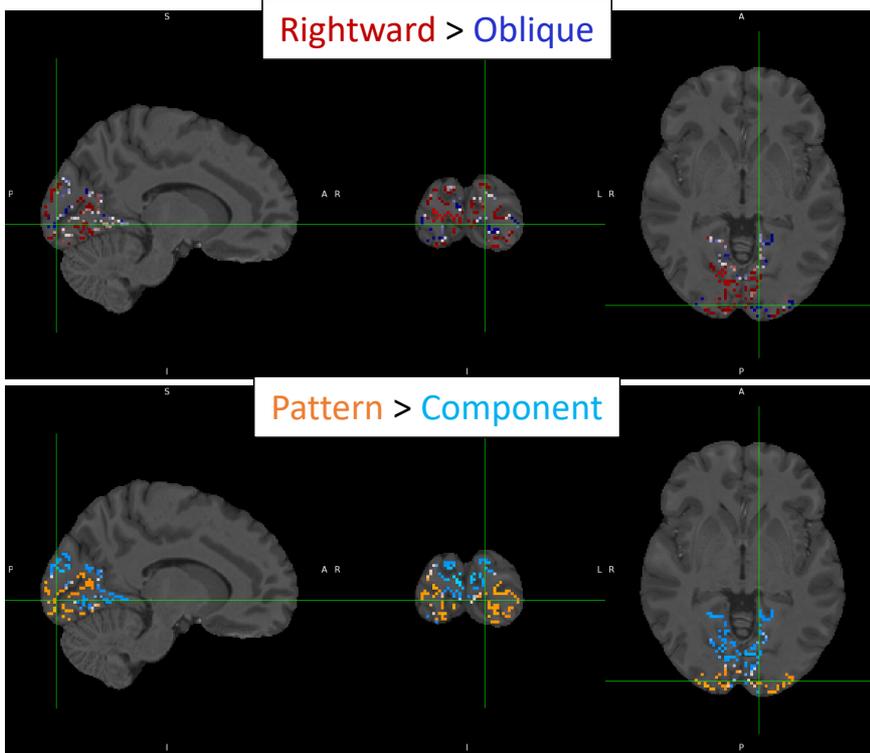


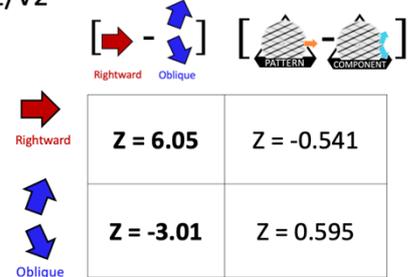
Figure S1.31: Direction-selectivity and Perceptual decision maps in subject 31
Localizer: In V1 (not in V2), Rightward > Oblique contrast estimates showed predominant response to Rightward direction (red).
Model failed in detecting Oblique-selective (blue) activity ($p < 0.01$ uncorr).
In hMT+ however, the reverse tendency was observed with a supremacy of Oblique response.
Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
This subject was discarded from the group-level analyses because of the undetected subdomains.

S32

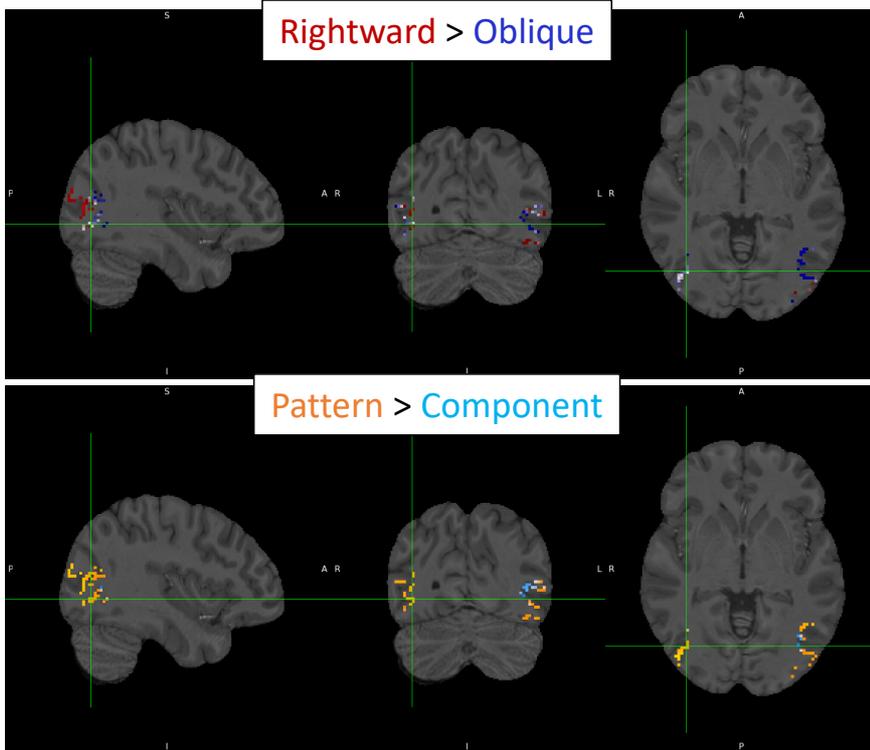
V1/V2



V1/V2



hMT+



hMT+

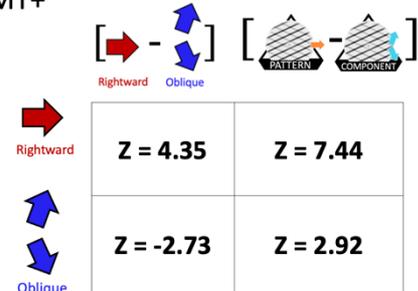
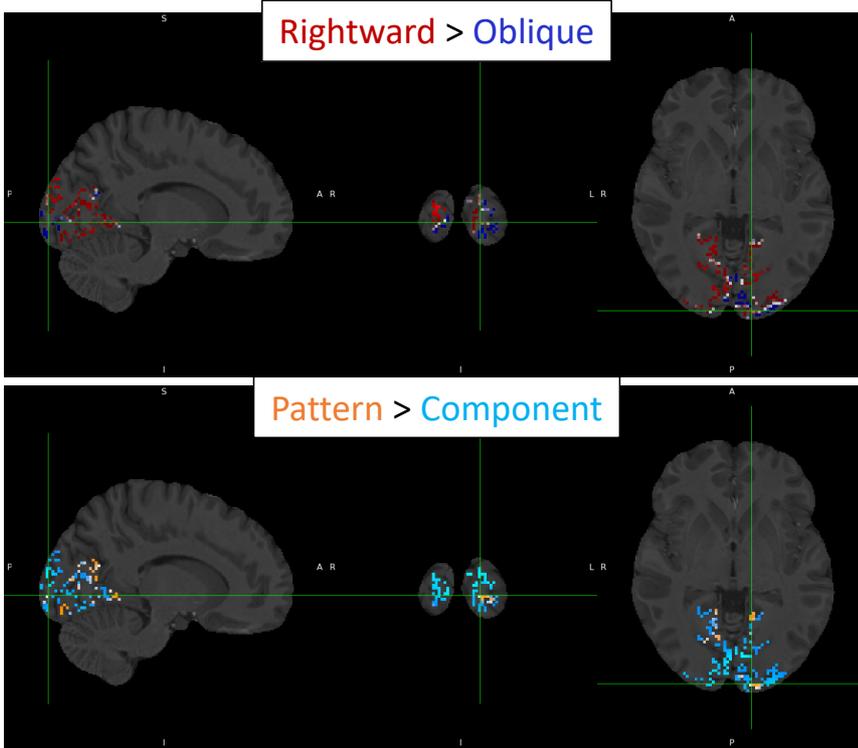
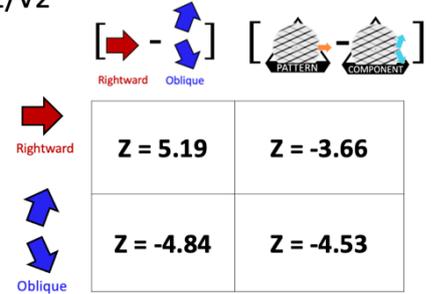


Figure S1.32: Direction-selectivity and Perceptual decision maps in subject 32
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

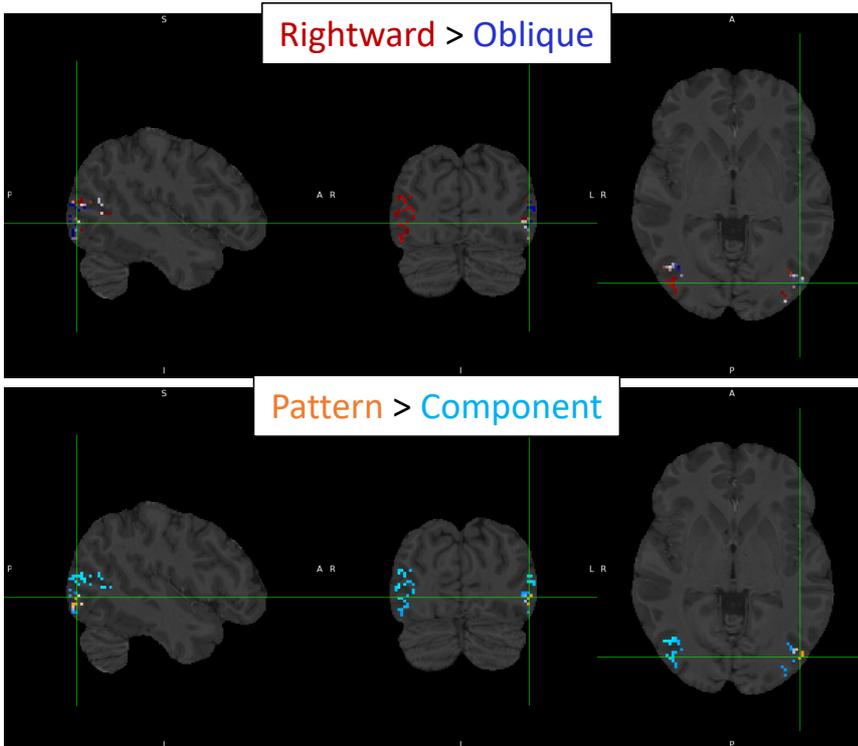
V1/V2



V1/V2



hMT+



hMT+

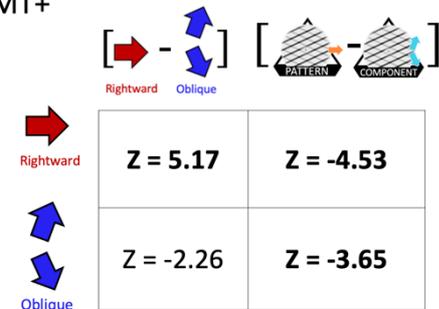


Figure S1.33: Direction-selectivity and Perceptual decision maps in subject 33

Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.05$ uncorr).

Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).

Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).

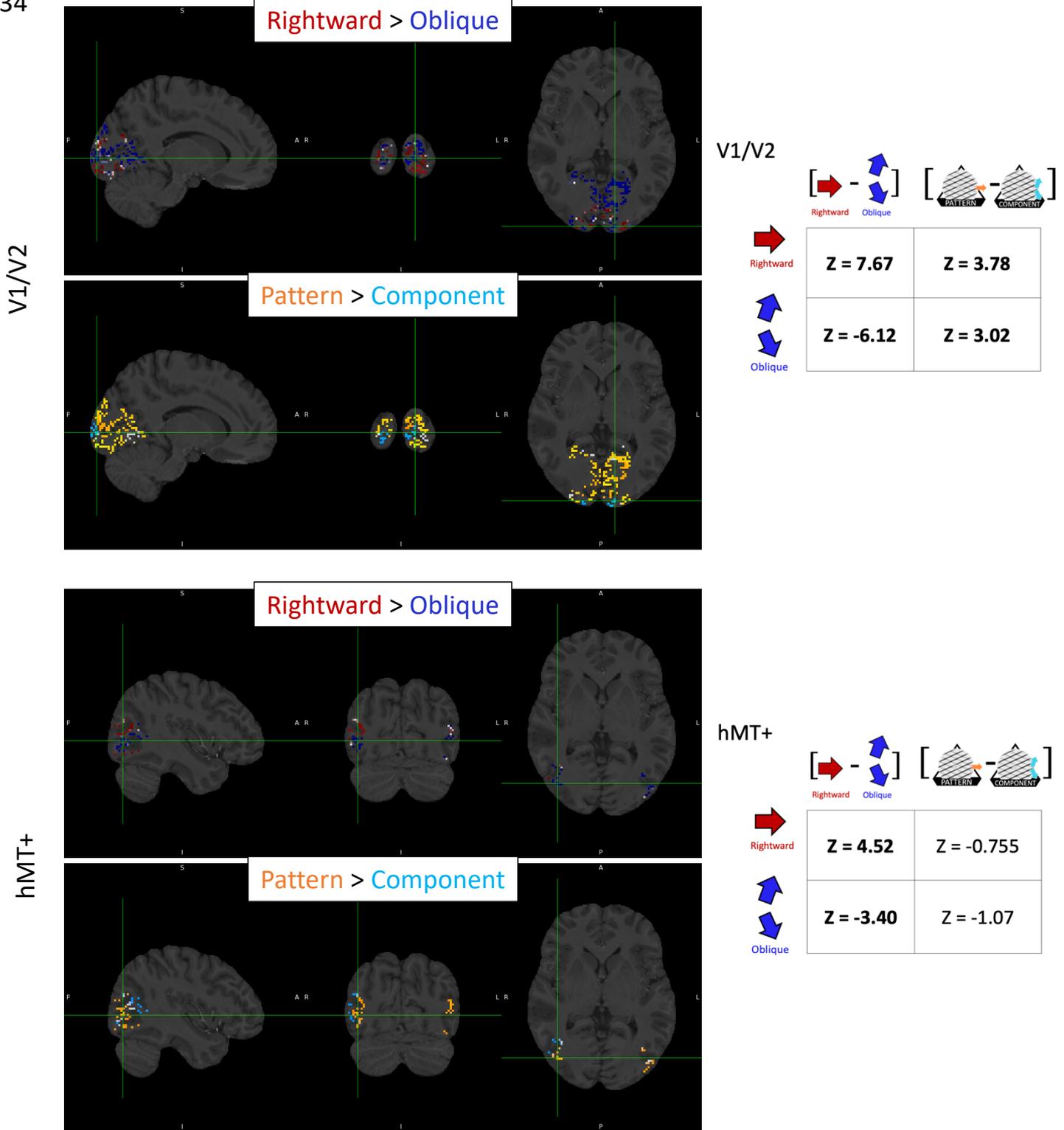


Figure SI.34: Direction-selectivity and Perceptual decision maps in subject 34
 Localizer: In V1/V2 and in hMT+, Rightward > Oblique contrast estimates showed Rightward- (red) and Oblique-selective (blue) activity ($p < 0.01$ uncorr).
 Bistable motion: Pattern > Component contrast resulted in Pattern- (orange) and Component-selective (cyan) activity ($p < 0.01$ uncorr).
 Tables report average z-value of the contrast estimates for the localized subdomains (25 voxels) showing highest Rightward or Oblique selectivity (in bold, z-value significantly different from zero: $p < 0.01$).