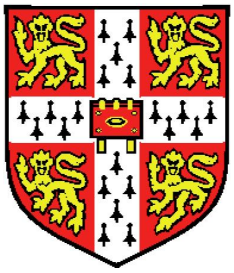
Neural computation of depth from binocular disparity



Nuno Reis Gonçalves

Sidney Sussex College

University of Cambridge

A thesis submitted for the degree of

Doctor of Philosophy

January 2018

To Ksenia and Ivan

Acknowledgements

The projects that form the basis of this thesis would not have been possible without the help of many. I would like to acknowledge them here.

Thanks to Rosa Panchuelo, Susan Francis, and especially Denis Schlüppeck for support with ultra-high field imaging at the University of Nottingham. Thank you also for your patience towards an overexcited first year graduate student.

I would also like to thank Jan Zimmermann, Valentin Kemper and Rainer Goebel for their help with ultra-high field imaging at the University of Maastricht. Thanks to Jan and Valentin too for caring deeply about the preparations for the experiments.

Thanks to my colleagues in the laboratory — they have all helped at some point. I am particularly grateful to Hiroshi Ban for his help during the first few months in the job; I learned a lot. To my colleague Caro Luft, now a lecturer at Queen Mary, University of London: thank you for looking after me.

I am extremely thankful to Andrew Welchman and Zoe Kourtzi for their advisory role, but also for the endless amount of opportunities and resources that they have always made available to me. I would also like to thank Ed Lalor, now an Associate Professor at the University of Rochester, for his support when I was a postgraduate student at Trinity College Dublin — my time in Ed's lab was a great experience and the springboard to an exciting doctoral degree.

I would like to thank the European Commission for funding my doctorate (FP7, Adaptive Brain Computations Initial Training Network) and the Erasmus Mundus programme that prepared me for this degree.

Most of all, I would like to thank my family — including my parents and my sister — for uninterrupted support and reassurance. To my wife Ksenia and my son Ivan: thanks for being part of this journey. This book is dedicated to you.

Preface

My time as a graduate student at Cambridge was beyond intellectually stimulating. In the laboratory, I worked with a variety of experimental techniques, such as magnetic resonance imaging, spectroscopy and psychophysics. I had the wonderful opportunity of combining experiments with modeling. But most importantly, I had the privilege of working on the fascinating topic of stereopsis. Some of the most remarkable minds of the last two millenia — Euclid, Da Vinci, Kepler, Newton, Descartes, and more — have been at some stage bewildered by binocular vision and stereopsis. They were mostly struggling with the geometry of binocular vision (i.e. how the light is captured by the left and right eyes and how might that relate to depth in the environment). This is now well known. What we don't know yet is how the brain uses the signals captured by the left and right eyes to estimate depth. This is the central point of the thesis.

Given that ultra-high field magnetic resonance imaging was not yet available in Cambridge, data acquisition for the experiments reported in Chapter 4 and 5 was performed in collaboration with the University of Nottingham and the University of Maastricht, respectively. The remaining contents of the thesis result from my own work, guided by my advisors Dr. Andrew Welchman and Prof. Zoe Kourtzi. I have not submitted any parts of the thesis for any other degree in the University of Cambridge or any other institution. The thesis does not exceed the word limit established by the Degree Committee for the Faculty of Biology.

Abstract

Stereopsis is a par excellence demonstration of the computational power that neural systems can encapsulate. How is the brain capable of swiftly transforming a stream of binocular two-dimensional signals into a cohesive three-dimensional percept? Many brain regions have been implicated in stereoscopic processing, but their roles remain poorly understood. This dissertation focuses on the contributions of primary and dorsomedial visual cortex. Using state-of-the-art machine learning techniques, we found that disparity encoding in primary visual cortex can be explained by shallow, feed-forward networks optimized to extract absolute depth from naturalistic images. These networks develop physiologically plausible receptive fields, and predict neural responses to highly unnatural stimuli commonly used in the laboratory. They do not necessarily relate to our experience of depth, but seem to act as a bottleneck for depth perception. Conversely, neural activity in downstream specialized areas is likely to be a more faithful correlate of depth perception. Using ultra-high field functional magnetic resonance imaging in humans, we revealed systematic and reproducible cortical organization for stereoscopic depth in dorsal visual areas V3A and V3B/KO. Within these regions, depth selectivity was inversely related to depth magnitude — a key characteristic of stereoscopic perception. Finally, we report evidence for a differential contribution of cortical layers in stereoscopic depth perception.

Contents

List of Figures

Chapter 1

Introduction

Animals with overlapping binocular vision extract three-dimensional information based on stereopsis. This thesis is about the neural computations (i.e. computations that might be carried out in the brain) that support this ability. I shall start with a basic introduction of stereoscopic perception, followed by its neural correlates. I will then introduce the most influential theories of neural computation of depth from binocular disparity. Although occasional links will be made to the field of computer vision, it is not my aim to focus on state-of-the-art algorithms for estimating depth from stereoscopic pairs.

For clarity, I will follow the terminology chosen by David Marr[1]. Specifically, I will use the term *disparity* to describe the angular difference between the projection of a point in three-dimensional space onto the left and right retinae. I will refer to the physical distance between that point and the observer as *distance*. Finally, the term *depth* will be used to describe the perceptual experience of distance.

1.1 Binocular vision and stereopsis

Behavioural relevance

The wonders of binocular vision have been appreciated by great intellectuals throughout the centuries, such as Ptolemy, Descartes and Newton[2]. Of interest was not only the so-called singleness of vision, but also the relationship between binocular vision and depth perception. For instance, Leonardo da Vinci noted that no single canvas

could elicit a realistic, vivid experience of depth because objects in 3-d space are often visible to one eye but not the other.

Ecology, not only aesthetics, attests the relevance of binocular vision and stereopsis. Early mammals have laterally positioned eyes, with only a minor portion of the visual field covered by the two eyes[3]. This provides them with a very generous coverage of the visual field—it maximizes the space of the environment that is captured by the retinae, while minimizing the amount of redundant (monocular) information. Such panoramic vision allows the detection of changes in the environment (e.g. an attack by a predator) even if they happen at the rear of the animal.

Conversely, early primates have front-facing eyes, which necessarily results in loss of panoramic vision and its important advantages. At such high cost, what benefits could front-facing eyes potentially bring? Several have been identified. For instance, animals with front-facing eyes may chase potential prey while capturing visual information with minimal optical aberration[3]. At the same time, because the retinae now capture overlapping visual information, binocular disparities can potentially be used to better estimate distances between different elements in the environment. This ability is useful for *(i)* identifying and capturing small prey[4]; *(ii)* grasping fine branches[5]; *(iii)* arboreal acrobatics[6]. Front-facing eyes could also be useful to break camouflage[7] and overcome occlusions that often occur in cluttered environments[8]. Importantly, binocular vision also supports less exotic behaviours, such as such as reading[9] or reaching and grasping objects in our everyday life[10].

Basic geometry of binocular vision

Our eyes are horizontally separated and, hence, each eye samples the visual world from a different vantage point. As a result, if we assume that an observer maintains constant the position of their eyes, the projections of a three-dimensional object onto the left and right retinae depend on the distance between the object and the observer. To illustrate this, let us consider a point object P in the three-dimensional space (Fig. 1.1). The observer fixates in a point F, which projects to the corresponding points F_L and F_R in the left and right retinae, respectively. However, the point P projects to non-corresponding points P_L and P_R in the retinae, depending on the distance between the point P and the observer. Absolute disparity is defined as the difference

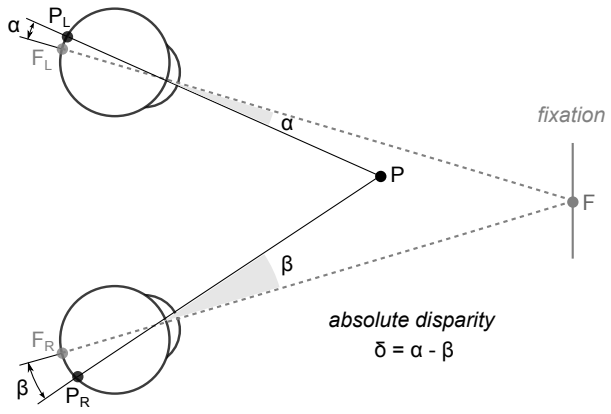


Figure 1.1: Basic geometry of stereopsis. Due to the horizontal separation of the eyes, points in 3-d space (e.g. point P) often project to non-corresponding location in the retina. The difference between the angles formed between these projections and the fixation point F is called binocular disparity.

in angular displacement between the projections of P and F: if we denote α and β as the angles between the projections of P and F onto the left and right retinae, then absolute disparity is then given by $\delta = \alpha - \beta$.

It is thus clear that absolute disparity depends on the position of P. However, it is important to notice that absolute disparity is also a function of the position of F (i.e. absolute disparity depends on vergence). Therefore, absolute disparity alone cannot be used to recover the distance between the point object and the observer — it can only tell us about the distance between P and F. To arrive at the absolute distance between P and the observer, one needs to know the distance between the observer and the point at which fixation is maintained, F. Vergence and accommodation signals could be used for this purpose.

The set of 3-D points that project to corresponding locations in the left and right retinae forms a surface known as the horopter. Since $\alpha = \beta$ for any pair of corresponding retinal projections, all points in the horopter thus have zero disparity. In the two-dimensional case (i.e. ignoring the vertical dimension), a good approximation to the horopter is the Vieth-Müller circle (Fig. 1.2). In this arrangement, points along the circle are considered to have zero disparity. Non-zero disparities are commonly classified as crossed or uncrossed. Crossed disparities are associated with temporal retinal displacements, while uncrossed disparities are associated with nasal retinal displacements. Objects with crossed disparity are therefore closer to the observer in relation to the fixation point, while objects with uncrossed disparity are

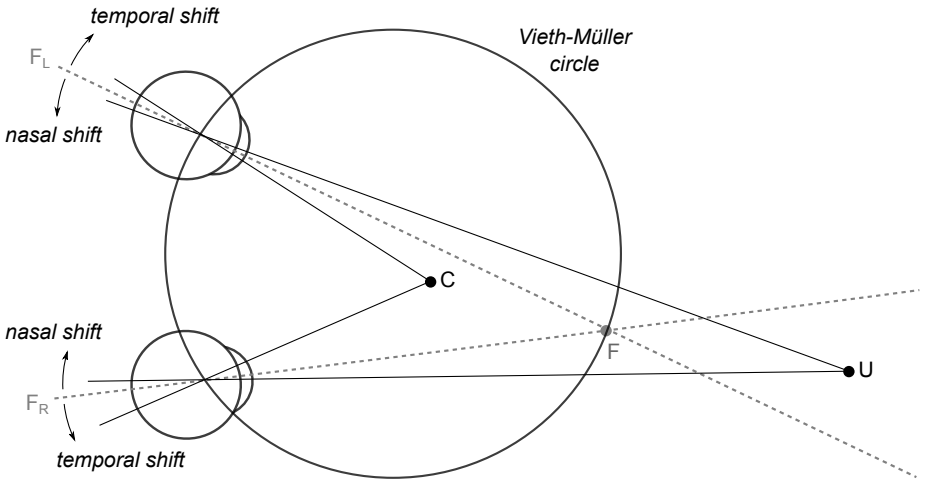


Figure 1.2: Disparity classification and the Vieth-Müller circle. Adapted from[2].

further away. Although the terms crossed and uncrossed disparity are usually used to denote the disparity produced by near and far objects, it is worth noting that it is not always the case that objects closer to the observer produce crossed disparities, nor that objects further away always produce uncrossed disparities. Strictly, the disparity associated to an object within the Vieth-Müller circle is termed convergent disparity, while the disparity produced by an object outside the Vieth-Müller circle is called divergent disparity[2].

Based on figure 1.2, we can also introduce the concept of relative disparity between any two points. The relative disparity between C and U is simply defined as the difference between their absolute disparities, $\delta_{CU} = \delta_C - \delta_U$. Since $\delta = \alpha - \beta$ (Fig. 1.1), the expression for relative disparity can be written as $\delta_{CU} = (\alpha_C - \alpha_U) - (\beta_C - \beta_U)$. Here, the absolute distance to the fixation point F is no longer determinant in the measure of disparity.

Disparity edges and occlusion

As Leonardo da Vinci observed, opaque objects in the foreground occlude certain regions of the background depending on the visual axis of each eye. For instance, let us consider a central fronto-parallel opaque surface closer to the observer in relation to a particular surround (Fig. 1.3a), in which the occluded regions of the background in each eye may be common or exclusive. The areas occluded in both retinal images

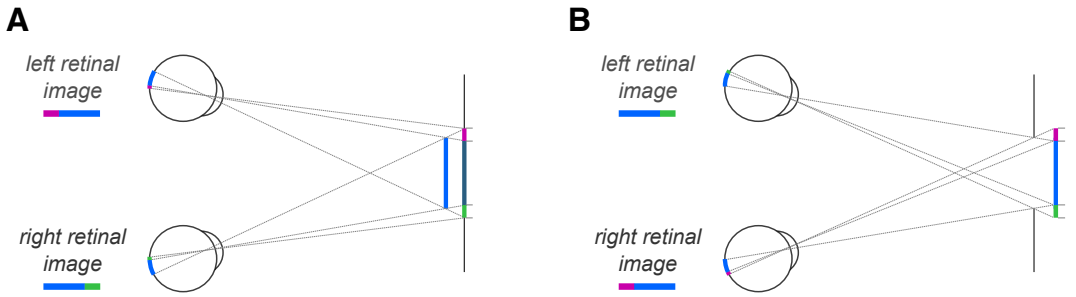


Figure 1.3: Geometry of stereoscopic occlusions. Monocular regions (green and pink) are exclusively captured by the left and right eyes. The position of the monocular regions in relation to the target depends on eye-of-origin information, and is diagnostic of depth (compare A and B).

are known as binocular occlusions, and are naturally not visible to the observer (Fig. 1.3a, dark turquoise area). Conversely, regions that are exclusively occluded in one of the retinal images are known as monocular occlusions (Fig. 1.3a, cyan and pink areas). Because monocular occluded regions, also known as *unpaired* regions, are not present in both eyes, there is no possible binocular correspondence between them. Hence, the classical definition of binocular disparity cannot be applied to visual elements in these regions. However, the retinal position of monocular occluded areas in relation to the central target (Fig. 1.3, blue) would be reversed if the surface was located further away from the observer in relation to the surround (Fig. 1.3b). It thus follows that the presence of unpaired features is a stereoscopic cue to depth. Nakayama and Shimojo[11] intuitively explain the problem at hand: the pattern of unpaired regions in the retinal images is associated with a *depth constraint zone*—due to the geometry of occlusions, a given unpaired feature is a projection of an object that can only be located in a particular region in space defined by visibility lines.

In this section, I have explained how differences between the left and right retinal images relate to distance between objects in 3-d space. In the next section, I will review behavioural observations which show that humans and other mammals make use of interocular image differences for depth perception.

Stereoscopic depth perception

We have previously seen how the horizontal separation of the eyes gives rise to binocular disparities. But can humans estimate depth based on such disparities? Charles Wheatstone provided the first evidence in support of this idea: he reported that if the

two-dimensional perspective views of a solid object (as seen from the vantage point of each eye) were presented separately to the left and right eyes, observers experienced the three-dimensional configuration of that object[12]. Whether binocular disparity alone could support depth perception was unknown, because Wheatstone's experiments lacked obviation of other cues to depth. Later, Heinrich Dove reported that stereopsis occurred for very brief stimulation periods before any eye movements could occur[13,14], suggesting that humans do not require vergence and accommodation cues to extract depth information from binocular disparity. Although initially in disagreement, Franciscus Donders later acknowledged that changes in vergence were not necessary in order to extract depth information from binocular disparity[15].

Apart from vergence and accommodation, Wheatstone's experiments also lacked obviation of other cues to depth. As Wheatstone defined it, the stimuli used in the experiment were the perspective views of a real object[12]. Therefore, perspective alone could be use as a cue to depth. Evident in Wheatstone's report[12] (cf. Figure N) are also additional signals that could be used for depth estimation, such as shading, occlusion and relative size cues. Thus, one question remained to be answered: are binocular disparities sufficient for depth perception?

A definitive answer to this question was given by Bela Julesz with the invention of the random-dot stereogram (RDS)[16]. This type of stimulus consisted of computer generated random patterns of black and white picture elements (pixels). Julesz generated stereoscopic pairs by displacing a given region of the stimulus in one eye in relation to the other. When binocularly fused, the displaced region of the stimulus is perceived closer or further away in relation to the remaining region of the stereogram (Fig. 1.4). Thus, the random-dot stereogram provided means of rendering three-dimensional structures by manipulating binocular disparity in isolation, without other visual cues to depth (e.g. perspective, occlusion, shading).

Stereopsis without correspondence

In his pioneering investigations, Julesz noted that depth percepts can emerge even in the absence of perfect correspondence between the left and right images[16]. In fact, even stereo pairs composed of very dissimilar features can effectively elicit a depth percept[17–19], provided they have overlapping spectral content[19]. Stereopsis is also possible when a single feature is presented monocularly[20,21]. Intuitively, these obser-

Figure 1.4: Computer generated random-dot stereogram, invented by Bela Julesz. Adapted from [16].

vations do not seem compatible with the detection of binocular disparities between corresponding features.

The study of stereopsis without correspondence is important because unpaired features occur frequently during natural stereoscopic viewing. For instance, an object in the foreground often causes occlusion of regions of the background in one eye, but not in the other (Fig. 1.3). As a result, certain image features visible to one eye will often lack a correspondent in the fellow eye (i.e. they are unpaired).

If stereopsis relies exclusively in positional differences between corresponding regions, then unpaired binocular features should be irrelevant or even detrimental for depth perception. Surprisingly, humans appear to take advantage of the presence of unpaired features [11,22–24]. In certain circumstances, perception of depth in stereograms containing unpaired features emerges faster than in stereograms that lack them [22]. Unpaired features that agree with occlusion geometry (Fig. 1.3, see caption) escape rivalry and, as expected, are perceived as part of the background surface [24]. So long as they are ecological valid, unpaired features can be qualitatively and quantitatively used for depth perception, at least over a local range of up to 25-40 arcmin [11]. Due to Leonardo da Vinci's early observations on the relationship between occlusion and depth, stereopsis based on unpaired features became known as *da Vinci stereopsis* [11].

While it is accepted that unpaired information affects stereoscopic perception, a debate persists on whether conventional and *da Vinci* stereopsis are supported by common mechanisms. Although it is unlikely that they rely exactly on the same processes [25,26], behavioural studies suggest that there is at least a fair amount of overlap. First, Pianta and colleagues observed cross-adaptation between the two types of stereopsis, and depth discrimination thresholds are strikingly similar [27]. Second,

depth from monocular occlusions can be biased by conventional disparity signals[28]. Finally, depth perception elicited by brief monocular stimulation[20] can in theory be supported by a matching-to-fovea process[21].

While stereopsis without correspondence is yet to be understood at the neural and perceptual levels[29], it is clear that (i) lack of binocular correspondence occurs routinely under natural stereoscopic viewing[11,30–32], and (ii) unpaired features can be perceived in depth[11,24].

Anticorrelated stereograms

Perhaps the most enigmatic observations on stereopsis relate to anticorrelated stereograms — stimuli in which elements are paired with inverted polarity across the half images (i.e. a bright element in one eye is paired with a dark element in the other eye, and vice-versa). Helmholtz had noticed that depth could be experienced in anticorrelated stereograms composed of thin lines, and the same holds for stereograms composed of bars[33]. However, less clear observations have been made on anticorrelated random-dot stereograms[33–37]. For very low dot densities, observers typically report a depth percept consistent with the binocular disparity imposed in the random-dot pattern[33,34], while for higher dot densities observers are usually unable to perceive depth[16,37]. To further complicate the matter, some observers seem to perceive reversed depth in anticorrelated stereograms[35,36,38], but only for very specific stimulus conditions (e.g. large disparity magnitudes[36]). As a result, the mechanisms that support depth perception based on anticorrelated stimuli remain elusive.

1.2 Neural basis of stereopsis

Understanding the relationship between neural activity and perception is a primary goal of sensory neuroscience. The last six decades have brought astounding progress to our understanding of how neurons in the brain encode information about the external world (i.e. neural encoding), and how these representations can be used to support perception (i.e. neural decoding).

Characterizing what features drive individual neurons is an important step towards understanding neural encoding. With great success, neurophysiologists have

been able to estimate spatial variations in brightness that best elicit neural responses in the retina[39–41] and in primary visual cortex[42–44], generating insights into the mechanisms by which contrast and orientation are encoded in the brain. While neurons selective for binocular disparity have been found more than five decades ago[45–47], understanding encoding of binocular disparity has proven somewhat challenging.

In this chapter, I will briefly introduce the visual pathways and how they convey binocular information. I will then move on to describing disparity selectivity in early visual areas, and from there segue to higher cortical regions. I shall focus on non-human primates due to the evolutionary proximity to humans and the body of evidence available, briefly mentioning other species when relevant. However, I note that disparity selective neurons have been found across many other animal models, such as cat[45], owl[48] and mouse[49].

Visual pathways and binocular information

Visual processing is hierarchically organized[50] (Fig. 1.5). Visual information captured by the retina propagates via the lateral geniculate nucleus (LGN) primarily to primary visual cortex (V1), which in turn projects to a network of cortical areas organized according to a distributed hierarchy. Many of these areas are involved in stereopsis (Fig. 1.5B, pink labels), but their precise contribution to stereoscopic perception remains unknown.

A prerequisite for encoding binocular disparity is to have access to the information captured by both eyes. Therefore, neurons that encode binocular disparity must be modulated by stimulation of either eye. In the retinae such neurons do not exist, and they are relatively rare in the LGN, where the majority of the fibers remain segregated according to eye-of-origin. Up to this point, individual neurons are typically driven by stimulation of one eye, but not the other. Neural activity is then relayed from the LGN to V1 via axons terminating in layers 4C and 6 (a small proportion of cells, namely part of the koniocellular pathway, project to other layers)[52].

The signals propagated up to layer 4C are still largely segregated by eye-of-origin[44]. Once in V1, the propagation of signals becomes more complex due to the abundance of feed-forward, horizontal and feedback connections[52–54]. The major pathways can however be captured by a simple model both in the cat and the macaque—

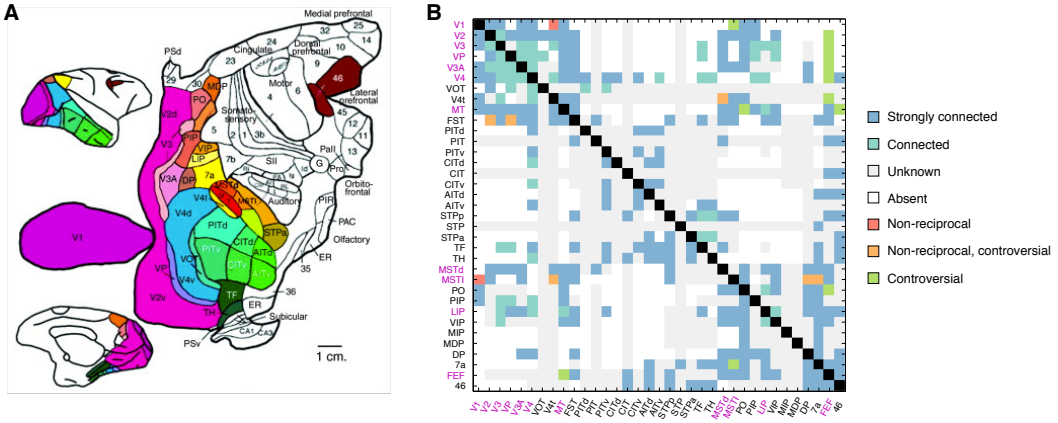


Figure 1.5: (A) map of cortical areas involved in vision in the macaque brain. (B) connectivity between the regions highlighted in color on panel A. Areas known to contain neurons selective for binocular disparity (summarized by Tsao and colleagues[51]) are labeled in pink (in addition to these, certain regions in inferotemporal and parietal cortices also contain neurons selective for binocular disparity, as I will discuss below). Adapted from[50]

intermediate layers receive geniculate input and project mainly to superficial (supragranular) layers, which in turn project to extrastriate cortex. Neurons in superficial layers also project to deeper (infragranular) layers, which provide feedback projections to intermediate and superficial layers[53]. In the macaque, layer 4C receives the majority of geniculate input (and weaker projections to layer 6). Layer 4C projects mostly to superficial layers 2-4B, which in turn project to layer 5. Deep layers 5 and 6 provide strong feedback to layer 2-4B and 4C, respectively[53].

Neurons in layer 4 of V1 are still modulated by stimulation of one eye alone[44] and are therefore not suitable for encoding binocular information. As we move towards supra- and infragranular layers the proportion of binocular neurons increases, with the highest proportion of binocular neurons being observed in supragranular layers 2/3[44,55]. Therefore, neurons in supragranular layers of V1 seem to be the first with access to the necessary primitives to encode binocular disparity. At later stages of the cortical hierarchy, the majority of the neurons are binocular[56,57].

Disparity encoding in primary visual cortex

Many neurons in primary visual cortex (V1) can be driven by stimulation of either the left or the right receptive fields[42,43,55], and stimulating both eyes produces stronger responses than stimulating one eye alone[42,43]. A possible mechanism for disparity

encoding could then rely on slightly disparate receptive field positions in the left and right eye. Indeed, many binocular neurons in cat V1 are driven by binocular disparate regions[45–47], thus suggesting that they could support stereopsis. Disparity selective neurons were later found in macaque V1[58,59], opening an avenue for more detailed investigations on how neurons in V1 may support stereopsis.

Based on the response to simple oriented stimuli, three major types of neurons were identified in V1: *simple*, *complex* and *hypercomplex* cells[43,44]. Simple cells are either excited or suppressed by presenting dark or light oriented stimuli in particular regions of their receptive fields. Their receptive fields have clearly defined excitatory (ON) and inhibitory (OFF) subregions. Complex cells, on the other hand, do not seem to be modulated by the precise position of the stimuli within their receptive fields[43,44]. Both simple and complex cells have been implicated in stereopsis. To my knowledge, a relationship between hypercomplex cells and stereopsis has not been explicitly established.

Binocular simple cells are thought to perform initial encoding of binocular disparity[60–64]. Consistent with Barlow's hypothesis[45], early investigations suggested that simple cells encode disparity via differences in *position* between the left and right receptive fields[65,66] (Fig. 1.6, left). However, subsequent studies revealed that the internal structure of the receptive field (i.e. the spatial arrangement of the ON and OFF subregions) varies considerably between the eyes[67–71]. Such interocular differences can be well accounted for by differences in the phase parameter of a Gabor function — hence the name *phase shifts*[71–73] (Fig. 1.6, right). Thus, disparity encoding using both interocular position and phase, also know as hybrid encoding, provides the best description of disparity encoding in binocular simple cells V1[68,69,71,73].

While binocular disparity modulates the firing rate of binocular simple cells, so does the overall stimulus position within the receptive field—even if disparity is kept constant[63,68,69]. That is, two stimuli with equal disparities presented in slightly different positions within the receptive field may cause very different responses. Therefore, the activity of an individual binocular simple cell is not sufficient to signal the presence of particular binocular disparity.

As opposed to binocular simple cells, binocular complex cells are largely invariant to position within their receptive field, and are therefore better candidates for disparity detectors[63]. The prevalent view was that complex cells inherit disparity

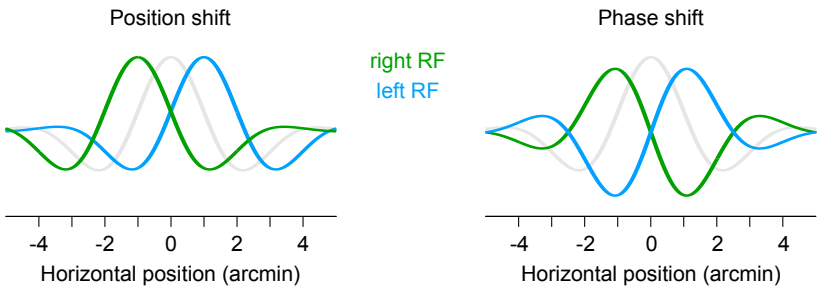


Figure 1.6: Left: receptive fields identical in structure but with different location in the two eyes can simply be related by a positional shift, and therefore they encode a positional disparity. Right: receptive fields with different internal structure in the two eyes can often be related by a shift in the phase parameter of a Gabor model. Blue and green lines represent the receptive field in the left and right eyes, respectively. Gray lines represent Gabors with zero position and zero phase shift.

selectivity by receiving excitatory input from multiple simple cells tuned to a particular preferred disparity[60,62–64]. It was also suggested that complex cells could integrate activity from other sources[74]—for instance, via direct geniculate inputs[75]—given the scarcity of binocular simple cells in the macaque brain[44].

There is now considerable evidence that individual complex cells receive inputs from neurons with a variety of disparity preferences. For instance, binocular complex cells can undergo adaptation by exposure to stimuli with non-preferred disparities[76]. They also seem to integrate activity over different positions, orientations and spatial frequencies[77–79]. Additionally, they receive suppressive input at opponent disparities[80–82]. These generalized inputs are thought to improve selectivity of complex cells for binocular disparity[79,81,82].

Can V1 complex cells then support stereoscopic perception? They are very diverse in their tuning properties, ranging from cells tuned to individual disparities, to cells tuned to a wide range of *near* or *far* disparities[58,73,83]. Thus, V1 complex cells provide a rich representation which could be used to support stereopsis. Although these cells are less selective than the perceptual system, a mechanism based on depth interpolation from a population of broadly tuned complex cells could in principle yield human-level stereoacuity thresholds[84,85]. Additionally, we know that complex cells are selective to disparity depicted in random-dot stereograms[83,86]. Together, these findings suggest that V1 complex cells could play an important role in supporting stereoscopic perception.

However, there are significant deviations between the activity of complex cells

and perceptual experience[87,88]. For instance, anticorrelated RDS do not typically elicit perception of depth (see section 1.1), but binocular complex cells still respond selectively to binocular disparities in these patterns[87]. For many binocular complex cells, responses to disparity in anticorrelated stereograms are inverted and attenuated in relation to their responses to correlated stereograms[87,89]. If binocular complex cells directly supported perception, an inversion in perceived depth would be expected, but this is hardly observed experimentally[33,37]. Additionally, even when binocular images are correlated, binocular complex cells can signal disparities that do not match perception[88,90,91]. Therefore, the activity of disparity complex cells in V1 seems insufficient to support stereoscopic depth perception, although it may provide a basis for fast corrective vergence eye movements[92]. Extrastriate cortex is thus expected to contribute to stereopsis.

Disparity encoding in extrastriate cortex

Disparity selective neurons are found throughout many extrastriate areas typically involved in vision[93–95], such as V2[58,96], V3/V3A[97,98], MT[99,100], MST[101], V4[102–105], and inferior-temporal cortex[106–109]. In addition, neurons selective to binocular disparity are also found in anterior[110–112], lateral[113,114] and ventral[115] parietal areas, as well as in the frontal eye fields[116], potentially for supporting visuomotor control.

Contrary to V1, the activity of many neurons across different extrastriate areas is closely linked to stereoscopic perception. Choice related neural activity during depth judgments based on disparity has been found in V2[90,117,118], V4[102] and anterior intraparietal area[110]. Additionally, microstimulation and inactivation studies have established a causal link between neural activity and depth judgments in areas MT[99], V2/V3[119], V4[102] and inferior-temporal cortex[120]. Interestingly, responses to anticorrelated disparities are greatly attenuated in V4[121] and absent at the level of inferior-temporal cortex[107]—in stark contrast with neurons in V1[87], MT[122] and MST[123].

Another point of differentiation between disparity encoding in primary and extrastriate visual cortex is the extent to which neurons are selectivity to relative disparities (i.e. differences between binocular disparities of different elements, such as the center and the surround of a circular RDS)[88]. Neurons specialized for different forms of relative disparity can be found as early as in V2[124], and to a higher degree in down-

stream areas V4[103], MT[125], MST[126,127], and inferior-temporal cortex[108,109]. These areas also seem to be specialized for different types of relative disparity. For instance, neurons in V2 and V4 respond well to relative disparities in center-surround configurations[103,124], while neurons in MT and CIP signal relative disparity in slanted surfaces[125,128]. Specialization for relative disparity in areas V3 and V3A is controversial — while neuroimaging studies suggest that V3/V3A are correlates of depth perception based on relative changes in disparity[51,129,130], neurophysiological data suggests no selectivity for relative disparity, at least for center-surround stimulus configurations[98].

The precise contribution of extrastriate visual areas for disparity processing remains unknown. They may be arranged into parallel hierarchical streams with different functional roles[131]. For instance, lesions show that the parvocellular and magnocellular pathways subserve different roles in stereopsis: damage to the parvocellular pathway impairs fine stereopsis, with no impact on coarse depth judgments, while damage to the magnocellular pathway leaves fine stereopsis unaffected[132]. It is therefore possible that different aspects of stereopsis are segregated according to these pathways[131].

A similar, potentially related, segregation in stereoscopic processing may also exist across ventral and dorsal streams[95]. For instance, dorsal area MT may underlie fast, coarse depth perception[133,134], and does not support fine disparity discrimination[134]. Parietal areas also show preference for fast, coarse depth perception[112]. Conversely, areas along the ventral stream show sharp selectivity for fine relative disparities that compose complex 3D shapes[102,103,105,107,108,120]. This suggests that the ventral stream may extract fine disparities more suitable for perception of 3D shape, while the dorsal stream may process coarse disparities for visually guided action[95,135]. Within the dorsal stream, different pathways are likely to coexist: for instance, sensory signals can propagate to parietal cortex indirectly via V3A[136], while they are conveyed in parallel to MT via direct and indirect input from V1. This indirect pathway to MT, particularly via V2/V3, is thought to be important for depth judgments[119,137].

Cortical organization for binocular disparity

In many species, V1 neurons are systematically organized along the cortex, whereby they cluster in columns with similar eye dominance and preferred orientation[43,44,138–141].

While the precise role of cortical columns remains unclear[142]. it is possible that emergence of perceptual related activity depends the presence of columnar organization[91]. In V1, cortical clustering according to preferred disparity is only weak[143,144], and is possibly driven by a correlation between preferred disparity and eye dominance[58,143].

By contrast, some extrastriate areas display cortical organization for binocular disparity. The first signs of clear organization for binocular disparity are found in V2: disparity selective neurons are clustered in the thick stripes of primate V2[145], and seem organized according to disparity preference[96,146]. In the cat, cortical organization for binocular disparity has also been observed[147]. Further down the dorsal stream, columnar organization for binocular disparity is found in areas V3/V3A[98,148,149] and MT[100].

1.3 Theory and physiological computation

Theory of stereoscopic perception

Stereopsis operates under very general conditions. We perceive depth in highly structured images[12], but also in random-dot patterns which lack the texture and form statistics that characterize natural images[150]. Although stereopsis relies on positional differences between the left and right retinal images, binocular correspondence is not strictly necessary to elicit depth percepts[11,18–20]. Moreover, we achieve stable depth percepts although fusion and vergence are in constant interaction[92,151]. Elaborating theories of stereopsis that encompass these different aspects is challenging, but necessary to guide future experimentation.

Early theories of stereopsis emerged within the field of experimental psychology[152–156]. They had been developed around unexplained psychophysical observations, which could in a way tell us something about how the brain operates. Take, for instance, the problem of the primitives of stereopsis. What visual features does the brain use to estimate depth from binocular disparity? To answer this question, one could experimentally control the features available to the visual system, and then evaluate whether stereopsis is achieved[17–19].

An alternative approach pioneered by David Marr focused on computational goals instead (here estimating depth from disparity). The *Marrian* logic is to depart from

the computation that needs to be done to solve a particular task, identify constraints, and propose an algorithm to solve the computational problem. This approach is agnostic as to how the solution is implemented in the brain, and focuses on the nature of the computation that the system must perform[157,158]. The computation, rather than the precise circuitry implementing it, can then be the used to relate neural activity to perception[159]. While some concerns exist with respect to this approach[160], it stimulated a wide variety of theorists to propose quantitative models of stereopsis.

Let us consider the problem of computing a binocular disparity. To do so, it is necessary to determine the correspondence between visual features present in the left and right retinal images—this is known as *the stereo-correspondence problem*. This is a challenging computation because images may contain many self-similar elements: in random-dot stereograms, for instance, each pixel in one image can be matched with many similar pixels in the other image.

In looking for a solution to the stereo-correspondence problem, perhaps the first consideration pertains to the substrate of matching, also known as matching primitives. What are the elements for which correspondence is to be established? Object recognition is not necessary for stereopsis, as demonstrated by Julesz' random-dot stereogram[150], so stereo matching does not operate on object elements or parts. Instead, matching should occur at the level of relatively simple features. One possibility is that the visual system matches very low-level features, such as the brightness of simple dot elements[158,161]. Alternatively, the visual system may operate on so-called zero crossings—areas where contrast polarity shifts from positive to negative and vice-versa[162,163]—or more complex features such as average brightness or contours[17,18,164,165]. Increasing the complexity of the primitives has the advantage of reducing the difficulty of the stereo-correspondence problem, but may also mean that simple features are left unmatched[31,161,162].

Another important issue is how the brain goes from a set of matching possibilities to a coherent solution, known as global stereopsis. Again, this problem is fairly evident in random-dot stereograms: a variety of equally good matches is available locally, and the correct match can only be determined by considering global information[7]. To deal with this problem, many theories of stereopsis use disparity detectors that are densely interconnected across space and depth[7,151,158,166,167], thus ensuring global support. One potential problem with this approach is that it requires a very large

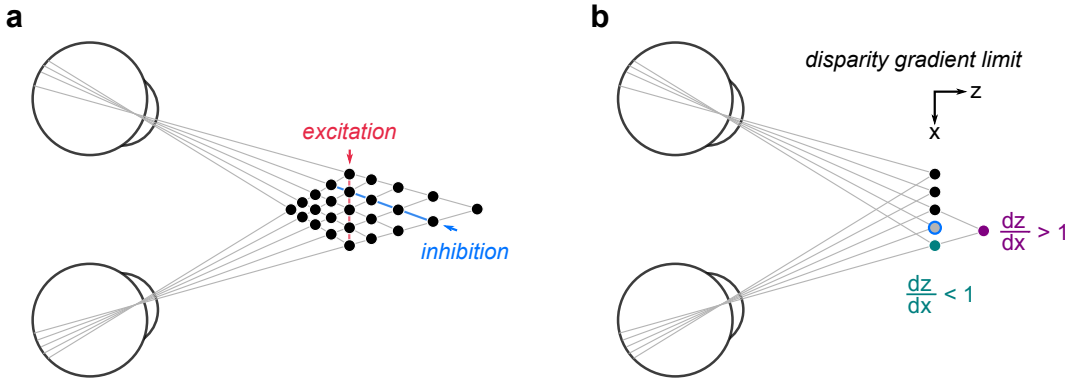


Figure 1.7: The stereo-correspondence problem and two strategies to overcome it. (A), continuity can be promoted by local excitation along the fronto-parallel direction. Inhibition along different lines of sight implement a uniqueness constraint. (B) elimination of a false match using the disparity gradient limit. In this example, one of two dots must be selected (turquoise or purple dot). The rate of change of disparity as a function of fronto-parallel displacement (with respect to the nearest match, blue dot) is used to discard false matches. Values greater than unity are considered violations of the disparity gradient limit, and such matches are therefore rejected. The criterion is based on perceptual observations[171].

number of connections. It is however possible to achieve global support with iterative local computations, which require a much smaller number of connections[168]. These theories suggest that the visual system may arrive at a global solution via local iterative computations and/or multi-scale interactions.

The stereo-correspondence problem can also be simplified by imposing constraints derived from ecological considerations. For instance, disparity in surfaces usually varies smoothly, so a constraint may be imposed that discards large local variations in disparity. In practice, this can be implemented by local excitation in the spatial domain[151,158,167,169] (Fig. 1.7a, red elements). A constraint may also be imposed in the rate of change of disparity across space[170] (Fig. 1.7b), or in local incoherence within different depth fields[166]. The stereo-correspondence problem can also be ameliorated by considering detectors tuned to different orientations and spatial frequencies[162]. Finally, inhibitory interactions between different disparity detectors can impose uniqueness on the disparity solution[151,158,167,169]. The solution is usually arrived at using cooperative interactions between different detectors[158,162].

These views make a central assumption that disparity can be extracted after matching corresponding features on the basis of their similarity. However, as I have alluded to above, it has been shown that human stereopsis operates on displays for which no

exact correspondence can be achieved[19,20], and that it can even be improved by introducing features that are only present in one half-image[11,22,24,31,172]. Thus, it has been suggested that more elaborate theories — that do not rely exclusively on matching — are needed to explain behavioural observations. One alternative theory is that the visual system explores dissimilarities between the half-images as a binocular cue to depth, possibly exploring three-dimensional ecological constraints[11,24,31]. This view has perhaps not received enough attention, but it is nevertheless plausible. Half-occluded regions are a cue to the presence of occluding contours, and therefore identifying such regions is informative for inferring the depth structure of a scene[31,32,173–175]. From a computational standpoint, an optimized system should be able to explore this dependency to estimate depth from binocular disparity. Recent theories propose that such occlusions are explicitly detected by a group of cells in the visual system[176], a prediction that is yet to be tested.

Finally, it is also important to consider the close relationship between stereopsis, vergence and accommodation. For instance, it is known that even very brief stereoscopic presentation elicits rapid, corrective vergence eye movements[92], and that disparity and accommodation dependent jointly affect eye-movements that facilitate binocular fusion[177]. Although vergence eye movements have been suggested to help in bringing large disparities into correspondence[162,163], few have attempted to relate stereopsis to the oculomotor system as whole[151].

The theoretical models above are either inspired or constrained by observations taken from early neurophysiological experiments. They consider general principles of cortical circuits, such as the existence of excitatory-inhibitory interactions[151,158,169] and the arrangement into cortical columns[167]. However, limited data on the properties of disparity selective neurons was available at that time. In the next section, I will review models of stereoscopic vision that aim to describe what the brain computes in order to estimate depth from binocular disparity.

Physiological models

The theoretical work hereto mentioned provides insights on the problems associated with stereoscopic perception. But what exactly does the visual system do to extract depth from binocular disparity? What are the neural computations involved? To

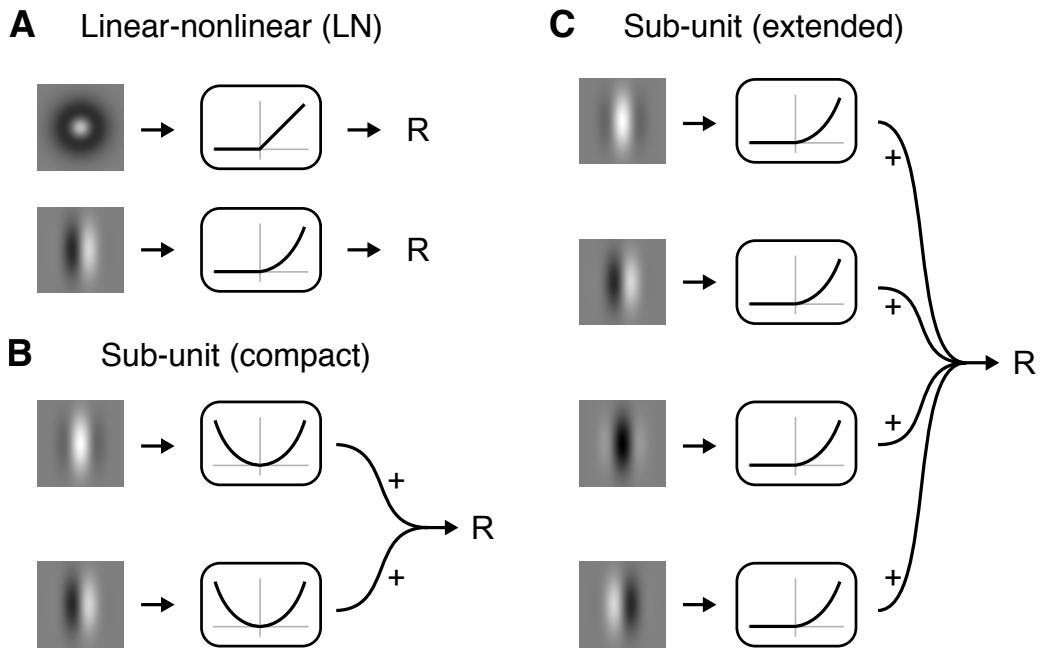


Figure 1.8: Simple models of physiological computation in early visual areas. (A) according to linear-nonlinear models, the response of a cell can be predicted by linear filtering by the receptive field (e.g., a centre-surround field as seen in the retina, or an oriented Gabor field as seen in V1) followed by a nonlinearity. (B, C) Subunit models combine the activity of simple cells (modeled an linear-nonlinear units) with receptive fields in quadrature-phase. The models represented in B and C are identical in their response, but they differ in their implementation. Panel B shows a compact representation where ON and OFF responses are computed by the same subunit, whereas panel C shows an extended representation where ON and OFF responses are computed by separate subunits.

answer this question, it is first important to characterize the computational abilities of neurons, i.e. the kind of computations that neurons are able to perform.

Significant advances in this quest have been made in the visual system, particularly in the retina[178], LGN[179] and V1[180]. For instance, simple cells in V1 (as X cells in the retina and LGN) behave quite linearly for a wide range of contrast energy, and their response can be well predicted on the basis of spatial summation over the receptive field followed by a rectification step that ensures non-negative firing rates[180] (Fig. 1.8A). Thus, these cells can be modeled as *linear-nonlinear* units, whereby the predicted activity of the cell is given by linear filtering of the stimulus by the cell's receptive field, followed by a static nonlinearity. To account for responses to high contrast stimuli, which deviate from the linear behaviour[181], subsequent divisive normalization is required[182].

Complex cells in V1 (as Y cells in the retina[178,183] and LGN[179]), on the other hand, deviate markedly from the linear-nonlinear model[184]. For instance, complex cells exhibit sustained responses to sweeping oriented lines, which effectively means that their response is not strongly affected by the position of the line within the receptive field. One possible mechanism to achieve position invariance relies on pooling the responses of simple cells (so-called subunits) with different spatial arrangements of ON-OFF receptive field regions[43,183,184]. The response of a complex cell can thus be modeled by pooling the responses of multiple linear-nonlinear units, each representing one simple cell (Fig. 1.8B,C). Subunit models are typically instantiated such that the receptive fields of the subunits are arranged in quadrature-phase, which ensures position invariance with the receptive field and equal responses to increments and decrements of light.

Following these building blocks of neural computation, Ohzawa and colleagues proposed a model that shaped the following two decades of investigations on stereopsis — the binocular energy model[63]. Using a subunit model, Ohzawa and colleagues proposed that disparity selective complex cells combine the output of four binocular simple cells with receptive fields in quadrature-phase (Fig. 1.9), with disparity selectivity emerging due to interocular differences between the left and right receptive fields. The original form of the binocular energy model includes a rectifying non-linearity at the output of the simple cells (Fig. 1.9), but modified versions of the model usually include additional nonlinearities (as I shall discuss later).

Let us define B as the set of responses of four simple cells in quadrature-phase, $B = \{b_0, b_{\pi/2}, b_{\pi}, b_{3\pi/2}\}$, where the subscripts denote the phase of the Gabor receptive fields. Formally, the response of a complex cell, c, is given by:

$$c = \sum_{b \in B} b. \tag{1.1}$$

In turn, the response of a simple cell is given by filtering the left and right images by the corresponding (i.e. left and right) receptive fields, summing the respective results, and applying a rectifying non-linearity. The response of a simple cell can thus be written as:

$$b = [l + r]_+^2, \tag{1.2}$$

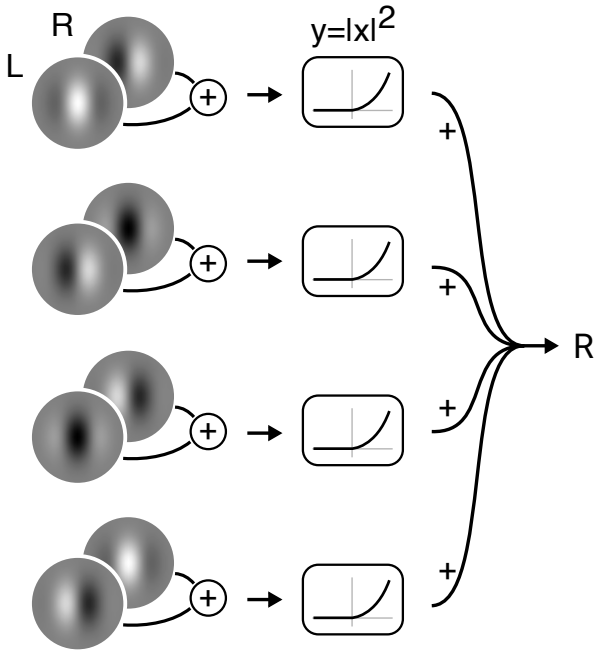


Figure 1.9: The architecture of the binocular energy model. For each subunit, the left and right images are convolved with the corresponding receptive fields. The filtering output for the left and right eyes are then summed (binocular summation) and the result is rectified. The rectified output of each subunit is then combined into an output complex cell. Modifications to the original energy model typically rely on additional nonlinearities at different stages of the model (e.g. before binocular summation).

where l and r are given by the dot product between the left/right input images and the left/right receptive fields, respectively, and $[.]_+$ denotes rectification. As mentioned earlier, the units of the energy model are arranged into two ON-OFF pairs. Thus, one can compute the aggregated response of two subunits in a pair by replacing the half-rectified squaring non-linearity by full-rectification. Thus, the response of a complex cell can be equivalently expressed by

$$c = (l_0 + r_0)^2 + (l_{\pi/2} + r_{\pi/2})^2, \tag{1.3}$$

where the subscripts again denote the phase of the Gabor receptive fields. By expanding the quadratic terms, it becomes clear that the response of a given complex cell depends not only on binocular interactions between the left and right images, but also on monocular information alone:

$$c = l_0^2 + r_0^2 + 2l_0 r_0 + l_{\pi/2}^2 + r_{\pi/2}^2 + 2l_{\pi/2} r_{\pi/2}. \tag{1.4}$$

Equation 1.4 is useful to understand how the binocular energy model works and what are its limitations. The response of the complex cell depends on six terms, only two of which are modulated by binocular information (the cross terms). Varying the disparity of the stimulus affects the binocular terms, but the quadratic terms are unaffected because they do not depend on binocular information.

Two important conclusions can be taken from this equation. First, the response of the complex cell depends on monocular contrast — for a given stereo image pair, increasing the contrast will cause an increase of the monocular responses l and r. Second, selectivity for disparity is supported by the cross-terms $2lr$. Thus, the complex cell's response is modulated by the covariance between the left and right inputs. As a result, by replacing one of the images by its negative, equation 1.4 explains why the binocular energy model produces inverted responses to anticorrelated stimuli — the monocular terms remain unaffected, but the sign of the cross-term is inverted.

Equation 1.4 demonstrates that the binocular energy model is too linear to account for the attenuation in response to anticorrelation[87]. A significant amount of

work in the last two decades has focused on modifying the energy model in order to account for this effect. A very simple way to explain the attenuation is the introduction of a threshold or non-linearity to the output of the complex cell[185,186]. However, this solution explains attenuation exclusively for tuned-excitatory cells and does not generalize to *near*, *far*, or *tuned-inhibitory* cells[186]. Another alternative is to introduce additional non-linearities before binocular combination, and modify the binocular summation stage of the model[186,187]. Finally, combining opponent excitatory-suppressive inputs from multiple subunits[82,188] can also lead to attenuated responses to anticorrelated stimuli.

Besides the difficulties in explaining properties of neurons, the classical binocular energy model also performs sub-optimally when it comes to computation. In the binocular energy model, the computation of depth from disparity is usually based on a maximum energy criterium, whereby the preferred disparity of the complex cell that responds the most is taken to be the predicted disparity of the stimulus[186,189,190]. However, it has been shown that the binocular energy model does not recover the correct disparity consistently: it very often produces spurious energy peaks at incorrect disparities, usually referred to as false matches[191,192].

One possible mechanism to abolish false matches is to pool the output of the energy model across multiple orientations and spatial scales[191]. This is effective because false peaks in energy occur at different disparities across different orientations and spatial scales, so they are attenuated by pooling. A more complex algorithm to eliminate false matches relies on the energy responses of detectors with a wide range of position and phase shifts, which can disambiguate the correct disparity in the stimulus[192].

Although physiological models of stereopsis have been gravitating around the binocular energy model, it remains unclear whether or not disparity selective complex cells receive input from simple cells, as suggested by subunit models. In fact, based on neural recordings from macaque primary visual cortex, Livingstone and Tsao found no evidence in support of this idea[74]. An alternative possibility is that disparity selective complex cells receive direct inputs from alternating layers of the LGN, with non-linear binocular integration happening at the dendritic level[75]. Although this model departs from the mechanistic account provided by the energy model, the computations implemented are closely related—this model too relies on

a linear-nonlinear binocular summation followed by integration of the activities of four subunits (in this case LGN efferents). The repertoire of neurophsysiological observations available to date is not yet sufficiently large to definitively tease apart the mechanisms by which disparity selectivity arises in V1 complex cells.

1.4 Thesis Overview

In the following chapters, I shall address three aspects of neural stereoscopic processing that remain poorly understood. I will start by looking at the computational mechanisms that may underly early disparity processing in the brain. The goal is to identify and describe a mechanism optimized to extract depth from binocular disparity, ideally based on natural images, and that can reproduce relevant neurophysiological and psychophysical data.

An re-evaluation of the relationship between early disparity processing and perception will follow. The aim will be to understand to which extent can the properties of disparity selective neurons in V1 be critical for the perception of depth.

Next, I will turn to the characterization of higher brain regions that are thought to be specialized for binocular disparity in humans. Previous neuroimaging work suggests that areas in dorsomedial visual cortex are closely related to depth perception[51,129,193]. I will examine these areas in greater spatial detail (for neuroimaging standards) using ultra-high field (7 Tesla) functional magnetic resonance imaging, with the main objective of testing for evidence of specialized cortical organization.

I will end with an effort to examine the role of different cortical layers in stereopsis using layer-dependent functional magnetic resonance imaging with sub-millimeter resolution. The main goal is to establish the feasibility of this technique to look at stereoscopic processing in the human brain, laying out the foundation for future work on exploring the role of cortical layers in mediating interactions between the very many visual areas involved in stereopsis.

Chapter 2

'What not' detectors help the brain see in depth

This chapter reproduces the work associated with the following published manuscript:

Goncalves NG & Welchman AE. "What not" detectors help the brain see in depth. *Current Biology*, 27, 1403–1412. 2017.

2.1 Introduction

Geometry dictates that a three-dimensional (3D) object viewed from the two eyes will (i) project features to different positions on the two retinae, and (ii) render certain portions visible to only one eye due to occlusion at the object's contours[12]. Computational[7,158,194] and neurophysiological[195] investigations over the past fifty years have focused almost exclusively on positional differences (i), as partial-occlusions (ii) are regarded as excessively under-constrained. Under this intuitive approach, by registering the positional difference of the same feature in the two eyes (*binocular disparity*), the brain could triangulate to infer the object's 3D structure. Thus, while the genesis of binocular information lies in image *differences*, current understanding at the computational- and neural- levels stresses the centrality of identifying *similarities* between the eyes to extract depth.

Within this framework, the fundamental challenge of stereopsis is described as solving the 'correspondence problem'[7,158,194] whereby images of the same real-world

feature are matched between the eyes. This is problematic because of 'false matches' — i.e., correspondences that conflate signals originating from different locations in 3D space. The principal means of identifying corresponding features is to consider a range of potential disparities and select the offset that maximises similarity between the eyes. This is captured computationally by the peak local cross-correlation. How might this be achieved by the brain? Current understanding is provided by the Disparity Energy Model of V1 neurons[63,189,191] in which binocular simple cells with disparity preference, δ_{pref}, are combined by a complex cell preferring the same disparity (Fig. 2.1A). Using a population of cells with different δ_{pref}, the brain could select the most active neuron to estimate depth.

However, from the perspective of finding correct matches, it is puzzling that many V1 neurons sense different things in the two eyes[67,71,73]. In particular, while binocular neurons can have receptive fields offset in location (*position disparity*), they often have different receptive field profiles in the two eyes (*phase disparity*) (Fig. 2.1B). The surprising implication is that phase neurons respond maximally to images that do not relate to a single physical feature in the world[192]. What are such responses for?

Here we suggest that V1 neurons should be understood as using a coding strategy designed to reduce uncertainty about the depth of the viewed scene. This involves the brain using both similar and dissimilar image features to infer depth. We show that long-standing puzzles in binocular vision at the physiological- and perceptual- levels can be understood by mixing feature detection with *proscription*. Specifically, by sensing *dissimilar* features the brain gains valuable information that drives suppression of unlikely interpretations of the scene. Our approach explains challenges to the standard treatment of disparity (i), and importantly, also accounts for (ii) partial occlusions that have long evaded explanation because of their incompatibility with registering depth based on peak cross-correlation.

2.2 Results

We start by considering known properties of binocular neurons from a statistical perspective[196], to demonstrate that properties that have long seemed puzzling in fact suggest optimal coding. Position-disparity units (Fig. 2.1B, purple) are easily understood from the traditional perspective: a viewed object will project its features to different

locations on the two retinae, so a binocular unit could simply offset the receptive field location for the two eyes. Phase-disparity units (Fig. 2.1B, orange), by contrast, have a different receptive field structure in the two eyes. This means they respond best to stimulation that could not originate from a single physical feature in the world. We contrasted phase- and position- encoding by computing Shannon information[196] as a function of stimulus disparity (see Methods), where simple units were modeled as linear filters followed by a rectified squaring non-linearity[63]. Because of the larger change in firing of the phase units, they provide more information about the viewed stimulus than position units (Fig. 2.1C). Importantly, the peak information provided by a phase unit is not at the traditionally-labelled δ_{pref} (i.e., peak firing rate), meaning that the Disparity Energy Model's architecture (Fig. 2.1A) of collating signals from units with the same δ_{pref} is likely to be suboptimal. We then examined encoding in a small population of simple units with *position*, *phase* or *hybrid* receptive fields. We found that *hybrid* encoding (i.e. combined phase and position shifts: Fig. 2.1B) conveys more information than either pure phase or position encoding (Fig. 2.1D). This suggests that the abundance of hybrid selectivity in V1 neurons[67,71,73] may relate to optimal encoding.

To test the idea that V1 neurons are optimised to extract binocular information, we developed a model system shaped by exposure to natural images. We implemented a binocular neural network (BNN, Fig. 2.2A) consisting of a bank of linear filters followed by a rectifying non-linearity. These 'simple units' were then pooled and read out by an output layer ('complex units'). The binocular receptive fields and readout weights were optimised by supervised training on a near *vs.* far depth discrimination task using patches from natural images (Supplementary Figure 2.8). Thereafter, the BNN classified depth in novel images with high accuracy (A=99.23%).

Optimisation with natural images produces units that resemble neurons

The optimised structure of the BNN resembled known properties of simple and complex neurons in three main respects. First, simple units' receptive fields were approximated by Gabor functions (Fig. 2.2B) that exploit hybrid encoding (Fig. 2.2C; Supplementary Figure 2.9)[67,71,73] with physiologically-plausible spatial frequency band-

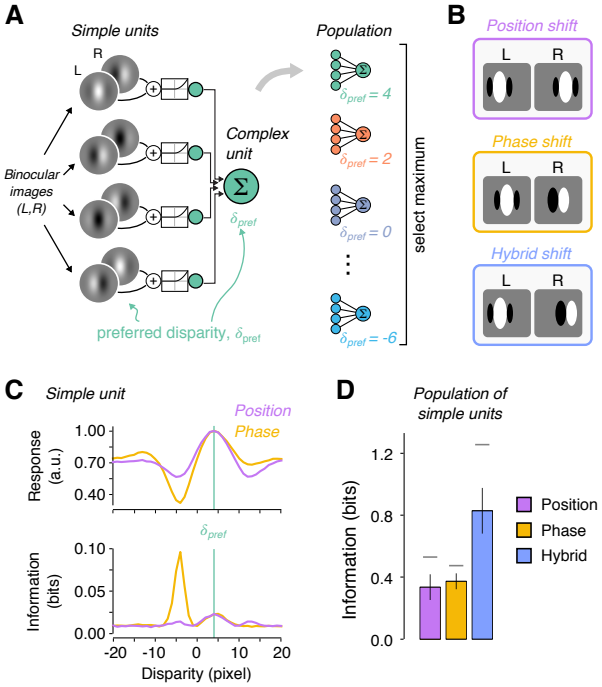


Figure 2.1: Disparity encoding and Shannon information. (**A**) The canonical Disparity Energy Model. Simple and complex units have the same preferred disparity, δ_{pref}. (**B**) Simple cells encode disparity using differences in receptive field: position (*position* disparity), structure (*phase* disparity), or both (*hybrid*). (**C**) Mean response of model simple units to 100,000 stereograms (top) and the corresponding Shannon information (bottom). Pink *vs.* yellow series contrast pure position vs. phase ($\pi/2$) encoding, both with δ_{pref}=4. Considering units between pure position and pure phase encoding produces a graceful morphing in the shapes of the curves. (**D**) Shannon information for a small population ($N = 5$) of simple units with position, phase or hybrid sensors. (Computing Shannon information for larger populations was computationally prohibitive). Error bars show SD over 1000 populations with randomly distributed phase and/or position shifts. Horizontal lines depict the upper limit on information determined by a population with uniformly spaced units.

widths (mean=2.3 octaves). Second, like V1 neurons, the BNN supported excellent decoding of depth in correlated random dot stereogram (cRDS) stimuli (Fig. 2.3A) (A=99.93%; $CI_{95\%}$=99.87%, 99.98%) that are traditionally used in the laboratory, despite being trained exclusively on natural images. Third, we tested the BNN with anticorrelated stimuli (aRDS) where disparity is depicted such that a dark dot in one eye corresponds to a bright dot in the other (Fig. 2.3A). Like V1 complex cells[63,87,89], disparity tuning was inverted and attenuated (Fig. 2.3B), causing systematic mispredictions of the stimulus depth (A=8.83%; $CI_{95\%}$=7.62%, 9.03%).

V1 complex cell attenuation for aRDS is not explained by the canonical energy

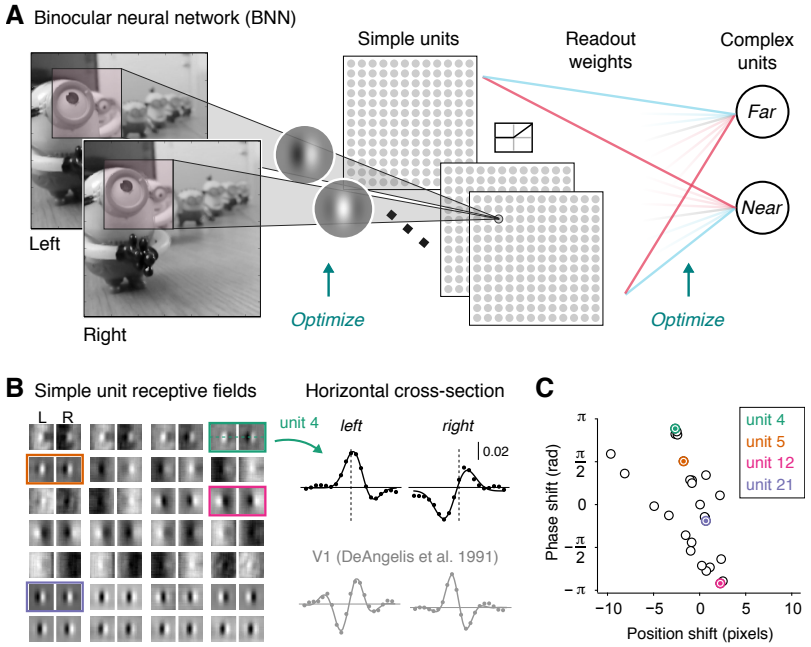


Figure 2.2: The Binocular Neural Network (BNN). (**A**) Network architecture: left and right images are filtered by simple units (28 binocular convolutional kernels), linearly rectified, and then readout by two output units. The form of the (i) receptive fields and (ii) readout weights was determined through back-propagation optimisation on *near vs. far* depth discrimination using patches from stereoscopic natural images[197]. The network learnt 21,254 parameters through exposure to 32,300 image pairs. (**B**) The BNN's optimised receptive fields resembled Gabor functions (mean explained variance by fitting Gabors to the 28 binocular receptive fields was $R^2 = 0.95$, $s.d. = 0.049$) and V1 receptive fields[67] (**C**) Summary of position and phase encoding by the simple units; representative units from (B) highlighted using colour. Note very few units show pure position or phase offsets. See also Supplementary Figure 2.8 and Supplementary Figure 2.9.

model, necessitating extensions that have posited additional non-linear stages[82,89,186,198]. However, the BNN naturally exhibited attenuation: by computing the ratio of responses to aRDS *vs.* cRDS, we found striking parallels to V1 neurons[87,89] (Fig. 2.3C). There was a divergence between the two comparison physiological datasets for low amplitude ratios, with our model closer to Samonds et al[89]. We speculate that this relates to the disparity selectivity of the sampled neurons: Cumming and Parker[87] recorded closer to the fovea where sharper disparity tuning functions might be expected. Accordingly, we observed greater attenuation (i.e., lower amplitude ratios) when the BNN was trained on multiway classifications (e.g., 7 output units, rather than 2) which produced more sharply tuned disparity responses (Supplementary Figure 2.10). Together, these results show that inversion and attenuation for anticorre-

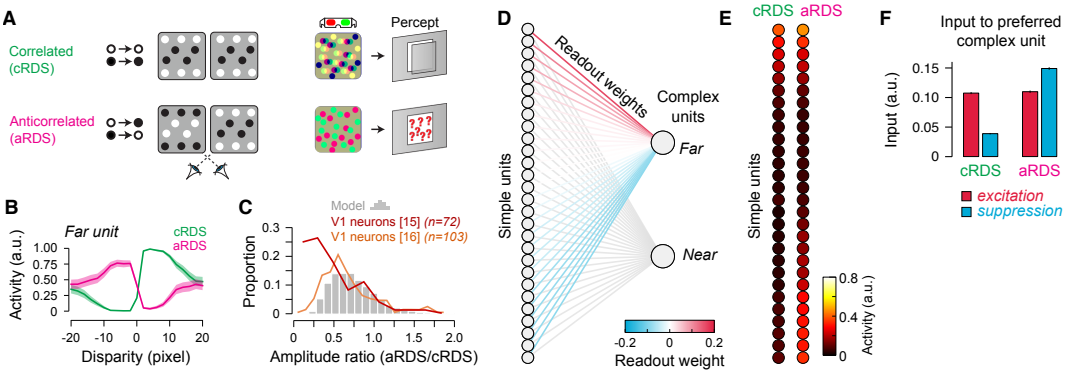


Figure 2.3: BNN response to correlated and anticorrelated random-dot stereograms. (**A**) Cartoons of correlated (cRDS, green) and anticorrelated (aRDS, pink) dot patterns with red-green anaglyph demonstrations. (**B**) Complex unit's disparity tuning curve for cRDS *vs.* aRDS; shaded area shows $CI_{95\%}$. (**C**) Distribution of amplitude ratios for cRDS *vs.* aRDS for the BNN (grey histogram; 5000 resamples), and macaque V1 neurons. Amplitude ratios were determined based on Gabor fits (average explained variance, $R^2 = 0.945$) (**D**) Representation of the weighted readout of the simple units. Units are ordered by their readout weight with *far*- preferred units at the top. (**E**) Mean activity for simple units in response to cRDS and aRDS. (**F**) Summary of excitatory (red) and suppressive (blue) drive to the output units for cRDS *vs.* aRDS. This represents the sum of the weighted simple unit activity split into the excitatory (positive weights) and suppressive (negative weights) components. Error bars (hardly visible) show $CI_{95\%}$. See also Supplementary Figure 2.10.

lation appear in a system optimised to process depth in natural images.

The traditional account of aRDS is that they simulate 'false matches' that the brain discards to solve the correspondence problem[107,121]. An alternative possibility, however, is that aRDS responses reflect a computational mechanism for extracting depth. To test this idea, we interrogated the BNN by ordering simple units by their readout weights (Fig. 2.3D) and then visualising the activity evoked by different stimulus types (Fig. 2.3E). The weighted-readout of simple unit activity defines the overall excitatory and suppressive drive to complex units in the network. We found that presenting aRDS led to a striking increase in the activity of the non-preferred simple units, while the activity of the preferred units was more-or-less unchanged. The consequence of this is that when this activity is readout it causes increased suppression at the preferred disparity (Fig. 2.3F). This changed the net drive to the complex unit from excitation to suppression (inversion), while the comparatively smaller difference between the excitatory and suppressive drives for aRDS produced a reduced amplitude (attenuation). Thus, attenuation and inversion can be understood based on changing the balance of excitation and suppression, without necessitating additional

processing stages.

To ensure that these parallels between the BNN and neurophysiology were not incidental, we tested whether he BNN produces outputs that are well-matched to the input stimuli. We used an optimisation procedure that started with random noise input images and iteratively adjusted the images such that the activity of a given complex unit was maximized (Fig. 2.4A). Following optimisation, the stimuli that best activated the complex units resembled a contrast edge horizontally translated between the eyes (Fig. 2.4B). Thus, the BNN is optimised for the translation of visual features that results from binocular viewing geometry[12]. Importantly, this is achieved using simple units that respond predominantly to different features in the two eyes (Fig. 2.2B), which are traditionally understood as 'false' matches (i.e., that do not correspond to the same physical real-world object). In other words, the BNN extracts depth structure without explicitly 'solving the correspondence problem'.

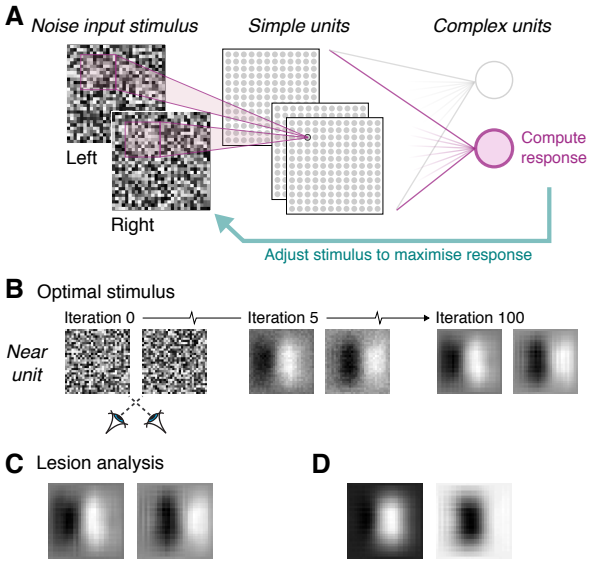


Figure 2.4: (**A**) Computing the optimal stimulus for a complex unit. Starting with random noise inputs, the algorithm computed the gradient of complex unit activity with respect to the input images. It iteratively adjusted the inputs to maximize the complex unit's activity. (**B**) Snapshots of three iterations during optimisation: a consistent On-Off pattern emerges in the left and right eyes, horizontally translated to match the preferred disparity of the unit. (**C**) This pattern remains when 'lesioning' the BNN of 25% of the simple units that use position encoding. (**D**) Removing highly-weighted hybrid units leads to input images that are unrealistic.

To strengthen this conclusion, we examined the consequences of 'lesioning' the BNN by removing 25% of its units. In particular, we removed units with near-zero

phase disparities (i.e., the seven units within $\pm\frac{\pi}{4}$ of zero phase offset) that are therefore best described as position disparity units that sense similar features in the two eyes. First, we considered decoding performance and found no effect on accuracy (A_{Pos}=99.97%, $CI_{95\%}$=99.92%,100%; p=.76; Supplementary Figure 2.9D). To situate this null result in the context of arbitrarily removing a quarter of the units, we also computed decoding performance when we randomly removed seven simple units. In this case, decoding performance dropped considerably (Supplementary Figure 2.9D), and there was only 3.8% chance of obtaining a value greater than A_{Pos}. This suggests that the pure position units contribute little to registering the binocular information by the BNN: they are given little weight so removing them has little effect relative to removing phase or hybrid units. Second, we computed the optimal stimulus for the lesioned BNN (Fig. 2.4C), finding little change relative to the uncompromised network. This null result was not inevitable: removing other simple units resulted in unrealistic images (Fig. 2.4D). Together, this indicates that the BNN does not critically depend on binocularly-matched features.

But how does the BNN extract depth using mismatches, and why should it respond to anticorrelated features? Under the traditional approach, this is a puzzle: a physical object at a given depth would not elicit a bright feature in one eye and a dark feature in the other. However, as we have seen, anticorrelation at the preferred disparity of a complex cell leads to strong suppression. This suggests a role for *proscription*: by sensing *dissimilar* features the brain extracts valuable information about unlikely interpretations.

The BNN accounts for unexplained perceptual results

If proscription has a perceptual correlate, then stereopsis should be affected by the availability of dissimilar features in the scene, an idea we now explore. First, seeing depth should be easier when there is more potential for anticorrelation at the *incorrect* disparity. This logic naturally explains a long-standing puzzle from the psychophysical literature[199,200] that demonstrated better judgments for stimuli comprising dark and bright dots (mixed polarity) compared to only dark or only bright dots (single polarity) (Fig. 2.5A). This result is difficult to accommodate within the Disparity Energy Model because correlation is largely unaffected by differences in the mean or amplitude of the input signals[200].

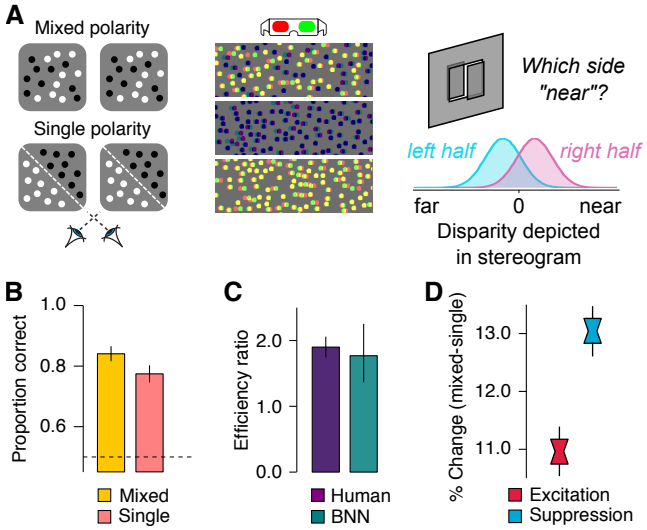


Figure 2.5: The BNN mirrors properties of human stereopsis. (**A**) Mixed vs. single polarity stereograms. Single polarity stereograms were either all dark, or all bright. The task was to discriminate the step arrangement of the stereogram. Anaglyphs designed for red filter over right eye. (**B**) Proportion of correct choices of the model after 1000 trials. (**C**) Efficiency ratio for mixed *vs.* single stimuli measured psychophysically [199] and for the BNN. (Note: the BNN was optimised on natural images, *not* random dot stereograms) (**D**) Difference between mixed and single stimuli in terms of the excitatory *vs.* suppressive drive to the non-preferred output unit. Error bars $CI_{95\%}$. See also Supplementary Figure 2.11.

We assessed the BNN's performance on mixed *vs.* single polarity stereograms (Fig. 2.5B), finding a benefit for mixed stimuli that was very closely matched to published psychophysical data [199,200] (Fig. 2.5C). What causes this improvement? As reviewed above, the network depends on the activity of the simple units moderated by readout weights. Presenting mixed *vs.* single polarity stimuli increases the simple unit activity, in turn changing the excitatory and suppressive drives to complex units. We found that mixed stimuli produce greater excitation for the preferred output unit and increased suppression to the non-preferred unit (Fig. 2.5D).

We carried out a number of controls to ensure that the BNN's performance was not artefactual. In particular, contrasting mixed *vs.* single polarity stereograms is complicated by low-level stimulus changes (e.g., overall luminance, or stimulus intensity range) that could act as covariates which underlie performance [200]. We directly manipulated covariate properties (Supplementary Figure 2.11), finding that the benefit for mixed stimuli persisted in all cases. We also tested the specificity of this result to the BNN's non-linearity [200]. Changing the nonlinearity to an unrectified squar-

ing operation did not change the result (Supplementary Figure 2.11). These controls indicate that the improvement for mixed stimuli generalises over perturbations of the stimuli and network architecture. These results suggest that performance improves for the mixed stimuli because of the opportunity to gain stronger evidence for the true disparity in conjunction with using mismatched features (i.e., dark-to-bright correspondences) as evidence against the incorrect disparity (i.e., proscription). This could be implemented *in vivo* using suppressive inputs to V1 neurons[81].

A second line of evidence in favour of proscription comes from considering situations regarded as too difficult for accounts of stereopsis based on peak correlation. Under natural viewing, certain features are visible to one eye but not the other (Fig. 2.6A). The brain exploits such unpaired elements, 'Da Vinci' stereopsis, to support depth perception[11,22]. However, these stimuli pose a severe challenge to traditional stereo-algorithms because there are no matching features[201]. We tested the BNN on a stimulus with unpaired features around a zero-disparity target (Fig. 2.6B). Because the target was not displaced in depth, there are no binocular corresponding features to compute the depth relationship. However, the BNN predicted the ordinal depth structure experienced by observers for the edge regions (Fig. 2.6B), and this result generalised to stimuli with different luminance configurations (Supplementary Figure 2.12). The BNN thus extracts critical signals that may provide the foundation for a full perceptual interpretation when used in conjunction with processes such as figure-ground segmentation at further stages of visual processing[130,176].

Finally, we tested the BNN on the classic 'wallpaper illusion'[202], in which periodic patterns yield ambiguous depth percepts. When disparity matches were ambiguous, the disparity-sign map did not identify a clear depth edge (Fig. 2.6C). However, by manipulating the background luminance to bias matching[31], we found that the BNN predicted the perceptual interpretation of the stereograms in the edge regions. This was achieved by changing the net excitatory-suppressive drive at the half-occluded regions, where disambiguation occurs (Fig. 2.6D). This is compatible with early processing of half-occluded edge regions in V1, providing an initial basis for subsequent depth interpolation supported by extrastriate cortex[203] or via recurrent connectivity within V1.

Together, these results indicate that, without being trained on such displays, the BNN's combination of detection and proscription provides a natural foundation for

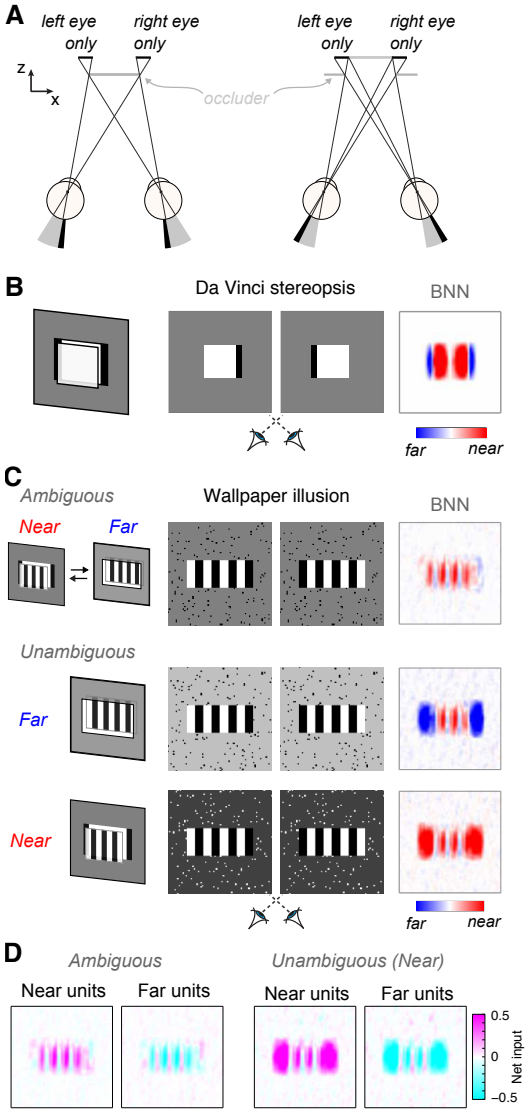


Figure 2.6: Ordinal depth prediction with ill-defined or ambiguous disparities. (**A**) Illustration of occlusion around the edges of objects. (**B**) 'Da Vinci' stereopsis. *Left*: Illustration of half-occlusions (black flanks) produced by viewing geometry; *Centre*: 'Da Vinci' stereograms for cross-eyed fusion; *Right*: depth map from the BNN. (**C**) Wallpaper illusion. *Top*: Ambiguous pattern. The vertical stripes can be matched by a nasal or temporal shift, making both *near* and *far* global matches valid. Cross-eyed fusion allows the reader to experience alternation. The BNN does not detect a clear depth. *Bottom*: Biasing perception by changing background luminance leads to a concomitant shift in the BNN's interpretation. (**D**) The net drive between excitation and suppression that underlies the shift in prediction, contrasting the ambiguous case and disambiguated cases. Note: for all these examples it is clear that the BNN has not 'reproduced' the percept: rather the network provides key signals that may provide the foundations for typical percepts. See also Supplementary Figure 2.12.

typical percepts. The simple units of the BNN exploit receptive fields that capture a continuum of similarities and differences between the binocular images, contrasting with the standard approach to binocular vision that emphasised the importance of correct matches. While individual units in the BNN are not specialised to identify the same feature in the two images, the aggregate readout activity classifies depth with high accuracy and complex units respond best to physically-realistic displacements of a single object.

Detection and proscription combine to facilitate sensory estimation

We have seen that the BNN generalises well from its training set and accounts for both neurophysiological and perceptual phenomena. However, the network's multiple parameters may act as a barrier to a detailed understanding of its operation. We therefore sought to explain the BNN's behaviour in theoretical terms by deriving a low-parameter closed-form model that captures its key characteristics. Our starting point was to observe that a low-dimensional rule relates the BNN's simple units and their readout: weights are proportional to the cross-correlogram between the (left and right) receptive fields (R=0.89; p<.001) (Supplementary Figure 2.13).

The key intuition behind this relationship is that receptive fields capturing a positive correlation at disparity δ_i (i.e. the lag of the cross-correlogram) should be read out by a complex unit with preferred disparity δ_i using a positive (i.e. excitatory) weight. Conversely, if the simple unit captures a negative correlation at disparity δ_i, the complex unit should read out its activity using a negative (suppressive) weight. In other words, the same simple units can be read out with detection or proscription to provide a population-based estimate of the depth of the viewed scene.

We show formally (Methods) that using weights determined by the cross-correlogram of the left and right receptive fields is optimal under reasonable assumptions, and propose a Binocular Likelihood Model captured by a simple equation,

$$\log L(\delta) = \sum_{i=1}^{N} r_i (W_L \star W_R)_i[\delta].$$

This relationship states that the activity of a complex unit that prefers a given disparity δ (expressed as a log likelihood, $L(\delta)$), is given by a weighted sum of simple unit

activity, r_i. The weights correspond to the cross-correlation, $(W_L \star W_R)_i$, between the left and right receptive fields of simple unit i at disparity δ (Fig. 2.7). To demonstrate the model, we implemented an instantiation that produces disparity tuning curves for correlated and anticorrelated RDS that closely resemble V1 complex cells (Supplementary Figure 2.14). This instantiation included a single spatial frequency channel, so the model does not require pooling across spatial scales to exhibit attenuation for aRDS. The model's key parameters are simply the receptive fields of the input units. This suggests that a fixed, stimulus-independent architecture explains key binocular phenomena, possibly without supervised learning.

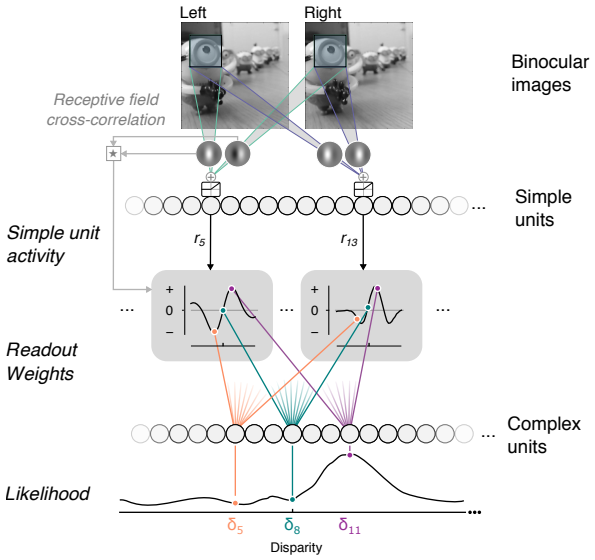


Figure 2.7: Binocular Likelihood Model. Input images are processed by a population of simple units that perform linear filtering followed by nonlinear rectification. The activity of a given simple unit (r_i) is readout by multiple complex units. A simple unit's readout weights vary over complex units, where the readout weight is defined by the cross-correlation of the simple unit's left and right receptive fields. The activity of the population of complex cells encodes the likelihood function for stimulus disparity. See also Supplementary Figure 2.13 and Supplementary Figure 2.14.

2.3 Discussion

Traditional understanding of stereopsis at the computational-, neural- and perceptual-levels has focused on the idea that peak correlation should be used to identify similar features and discard false matches. The logic underlying this approach is based on

inverting the geometry that maps objects at different locations in space onto different portions of the two retinae. However, here we show that envisaging neurons as units that match up the features of objects in the world fails to account for known properties of neurons, and overemphasises the role of similarity in a system whose fundamental benefit lies in differences between the images sensed by the two eyes.

We demonstrate that V1 neurons have properties ideally suited to extract binocular information, rather than simply searching for matching features. We formalise a Binocular Likelihood Model that provides a unifying account for previously puzzling properties of V1 neurons as well as perceptual phenomena that challenge the standard approach. This model highlights the interplay between feature detection and proscription for perceptual inference. This mix of evidence for and against likely interpretations may represent a general strategy for perceptual integration both within and between sensory modalities.

Understanding the functional role of sensory neurons

Understanding the coding strategies of sensory neurons represents a longstanding challenge. A historically pervasive idea is that sensory neurons act as 'feature detectors', signalling evidence for the occurrence of a particular feature in the environment[39,204]. For instance, orientation-selective neurons could indicate the presence of a particular tilted edge in a visual display[42]. It has long been recognised that natural images shape this selectivity[205,206], with neural responses optimised for efficient representation of the statistical regularities of the environment[207,208].

Here we take the approach of quantifying the information conveyed by early sensory neurons that are sensitive to binocular disparity using information analysis, and then by implementing a neural network optimised by exposure to natural images. This provides insight into the functional purposes of disparity representations at the neural and perceptual levels. Our findings on the utility of hybrid receptive fields for disparity encoding are consistent with work that used dimensionality reduction to estimate the optimal disparity filters[209]. In particular, our observation that hybrid units capture greater Shannon information is consistent with the idea that hybrid encoding maximises disparity estimation accuracy. Moreover, hybrid receptive fields are suggested to minimise the statistical redundancy of binocular responses[210–212], suggesting an additional factor driving the brain's use of hybrid units.

Understanding the encoding properties of the BNN

Previously it was suggested that phase encoding is used to sense 'impossible' stimuli. In particular, Read and Cumming[192] made an important proposal that key depth information is conveyed by positional disparities, with phase disparity used to select between alternative positional signals in cases of ambiguity. They suggested this would filter out 'false' matches, and thereby solve the correspondence problem. In contrast, our model is based on the combination of feature detection and proscription, rather than using mismatches as a veto. As we have shown, extracting depth structure can be achieved without units that register pure positional disparities: only 3/28 simple units responded to position offsets without phase offsets (Fig. 2.2B) and removing units with small phase offsets had little consequence on the performance of the network (Fig. 2.4C).

More generally, it is important to ask why the BNN, optimised by natural images, uses hybrid encoding for its simple units. The traditional exposition of binocular vision starts from the convenient geometry of how a small number of isolated points in the world project into the retinal images sensed by the two eyes. Models of binocular vision are typically built upon the logic of inverting this mapping based on establishing the 'correct' matches. However, the BNN suggests that the diet of early visual neurons consists almost entirely of mismatched features: the one 'true' set of correspondences between the two eyes is engulfed by a preponderance of mismatches.

When interpreting the properties of the BNN it is important to recall that the network learnt the relationship between specific inputs (i.e., one natural image set) and the optimisation objective (i.e., a particular discrimination task). Systematically changing either would change the learnt model. Nevertheless, the BNN generalised to a different stimulus set (random dot patterns) and had properties mirroring neurophysiology. It is interesting that the BNN's receptive fields are vertically-oriented. While this makes sense when capturing horizontal disparities, real V1 binocular neurons have varied orientation tuning preferences[67]. This difference may relate to the fact that the BNN is constrained to optimise one task (disparity discrimination) while V1 neurons are required to support many. It will be interesting to test how defining models for multiple objectives (e.g., estimating the orientation of features tilted in depth) affects encoding properties. For instance, future work might test whether

units become specialised for particular functions *vs.* develop joint-encoding characteristics. This might most straightforwardly be applied to proscriptive processing for motion estimation (given the strong computational similarities between disparity and motion[213]), but may also extend to other feature dimensions.

Relation to the disparity energy model

The disparity energy model[63,189,191] has long provided the foundation for understanding binocular vision. While modifications have been proposed to accommodate a number of electrophysiological observations[82,89,186], the basic architecture has remained unchanged. Moreover, the link between the implementation and the computational goal of estimating depth has been left obscure.

Here we developed an approach that exploits the same computational building blocks as the traditional model (i.e., linear filters for binocular summation followed by rectification). However, the BLM uses a weighted readout scheme, in which activity can be combined via excitatory or suppressive weights onto a population of complex cells. The main deviations from the traditional model are 1) the existence of multiple simple cell-like neurons, as opposed to the quadrature pairs originally proposed, 2) the incorporation of variable weights that can be suppressive, and 3) the complex unit's use of responses from simple units that do not have the same preferred disparity (because simple units convey information about multiple disparities). These characteristics are not part of the classical energy model, but strongly align with modifications suggested in light of neurophysiological evidence[77,78,80–82]. As we have shown, by using a model optimised to estimate depth, readout weights can be derived directly from the model's encoding properties. The fact that doing this reproduces properties of simple and complex cells measured *in vivo* suggests that the visual system has been optimized by similar constraints.

The role we demonstrate for proscription is consistent with evidence that binocular V1 neurons are modulated by excitatory and suppressive components[81]. That suppression lags behind excitation by $\sim$7 ms[80] suggests that it is initiated at very early stages of processing. In particular, the proscriptive registration of dissimilarities could drive suppression of unlikely depths via inhibitory interneurons. The necessity of an additional synapse (via interneurons) would impose a small temporal delay, but this delay is less than would be expected for extra-striate feedback. The BLM suggests

that the properties of suppressive inputs shape the inversion and attenuation of complex cell tuning curves for aRDS. Where suppressive input is strong, we expect a clear inversion of the tuning curve, but little attenuation. Conversely, where suppressive input is weak, such that excitation and suppression are nearly balanced, the tuning curve would be severely attenuated. In this case, the close balance between excitatory and suppressive inputs means that highly attenuated cells take longer to cross their firing threshold. This is consistent evidence from barn owls that longer onset latencies are associated with high attenuation[214].

Finally, the BLM predicts that anticorrelation masks the registration of a correlated disparity signal. Previous work pitted cRDS against aRDS to produce zero net correlation in the display. Participants can judge depth in such displays, leading to the suggestion of an additional mechanism separate from correlation[36]. In contrast, the BLM posits a single mechanism, and exploits anticorrelation to facilitate the interpretation of depth. We predict that the masking effects of anticorrelation are tuned (i.e., anticorrelated disparities are more suppressed than others) and that spatial limits on masking from anticorrelation are set by V1 complex cell receptive fields.

Relation to binocular rivalry

Our mechanistic account of the early stages of binocular vision suggests a natural link to work on binocular rivalry. Traditionally, the study of rivalry and stereopsis have been separate[215,216], although recent work suggested computational links between them[217]. Here we show that proscription is likely to be a key constituent of normal disparity processing. This suggests that stereopsis and rivalry sit along a spectrum of binocular responses mediated by inhibition. This is compatible with work on the perception of visual appearance[218] and suggests a link to GABA-mediated inhibition related to binocular rivalry. For instance, there is a strong association between human V1 GABA concentration (quantified by Magnetic Resonance Spectroscopy) and monocular percept duration[219]. Further, temporary monocular deprivation leads to reduced V1 GABA[220]. Therefore, it seems plausible that inhibitory mechanisms in V1 are related to processing binocular incongruence. It will be interesting to test how the mechanisms we propose are implemented physiologically, and whether these support a unifying axis between rivalry and stereopsis.

Relation to cue integration and multisensory processing

Finally, it is worth noting that neuronal tuning to properties that appear inconsistent with the physical structure of the world are not limited to binocular disparity. In particular, neurons can be tuned to the same or opposite features for different visual cues and/or between sensory modalities[221–223]. For instance, certain neurons in macaque area MSTd respond maximally to the same direction of motion when specified either by visual- or by vestibular- cues ('congruent'); while others ('incongruent'), have opposite direction preferences between modalities[223]. As with the discussion of phase disparity, 'incongruent' neurons are puzzling because they respond best to stimulation that could not be caused by a single physical object.

The inference framework we provide for binocular vision suggests an important role for neurons that encode proscriptive features. We hypothesise that a similar mechanism is used when combining different cues (e.g., disparity and texture) or sensory modalities (e.g., vision and touch). Specifically, neurons form a continuum of responses (ranging from 'congruent' to 'incongruent') analogous to 'hybrid' disparity encoding. These encoding neurons can be read out by a population of units that integrate signals from different cues. This can broadly be conceptualised as a type of causal inference based on explaining away[224] and links to suggestions about providing a mechanism for discounting irrelevant properties of viewed stimuli[225].

Conclusion

Early sensory neurons are broadly understood as optimised to capture the physical properties of the surrounding environment. Within this context, neural tuning to elements that do not relate to physical objects represents a significant puzzle. Using an optimal information framework, we demonstrate the importance of proscription: neural responses that provide evidence against interpretations incompatible with the physical causes of sensations. We demonstrate the role of these 'what not' responses in a neural network optimised to extract depth in natural images. We show that combining detection with proscription provides a unified account of key physiological and perceptual observations in 3D vision that are unexplained by traditional approaches. We capture the encoding and readout mechanisms in simple analytical form, and propose that marrying detection with proscription provides an effective coding strategy

for sensory estimation.

2.4 Methods

Information theoretic analysis

Individual simple units

We sought to formalise the idea that information encoded in the responses of binocular simple units is not restricted to the preferred disparity. To do so, we computed the Shannon information I between broadband stimuli s with varying disparity δ and simple unit responses R,

$$I(R, s_\delta) = \sum_i p(r_i|s_\delta) \log \frac{p(r_i|s_\delta)}{p(r_i)}, \tag{2.1}$$

where r_i denotes the firing rate of the simple unit. The resulting information indicates how well a particular disparity is encoded in the response of the simple unit. In this demonstration, the receptive fields were parameterized as two-dimensional (x, y), vertically oriented Gabor functions,

$$W(x, y) = e^{((x-x_0)^2+y^2)/2\sigma^2} \cos(2\pi f(x - x_0) + \phi), \tag{2.2}$$

where σ denotes the Gaussian envelope width, x_0 denotes the position, f the spatial frequency, and ϕ denotes the phase of the receptive field. To define the disparity encoded by the simple unit, we varied the phase and/or position, and kept the remaining parameters constant. Varying the position parameter introduces a simple translation in the receptive field, while varying the phase causes a change in the internal structure of the receptive field.

We computed the information carried by a simple unit with preferred disparity of 4 pixels defined by a either a position shift or a phase shift. For this simulation, the receptive field envelope, σ, was set to 5 pixels and the frequency, f, was set to 0.05 cycles/pixel. The stimulus set consisted of 100,000 uniform random dot images with disparities between —20 and 20 pixels. For both encoding mechanisms, we observed that individual simple units convey information about non-preferred disparities (Fig.

2.1C). This highlights that the activity of simple units selective for a particular disparity could contribute to the activity of complex units tuned to different disparities.

Population of simple units

In the previous section we examined information at the single unit level. Next, we demonstrate how much information is encoded across a small population of simple units ($N = 5$) with position, phase and hybrid disparity encoding. We used a small number of units for computational convenience, as the amount of memory required to store the full stimulus-response distribution increased exponentially with the number of units (simulating a population of 10 units, for instance, would require a prohibitive 80 gigabytes of RAM memory). An alternative to study information in larger neural populations would be to use other measures such as the linear Fisher Information – a quantity that is inversely related to discrimination thresholds, and that can be efficiently computed if responses follow a distribution of the exponential family with linear sufficient statistics[226]. However, we chose to use Shannon Information to avoid focusing on discrimination tasks and obviate further assumptions about the response distribution.

Although we are now working at the level of multiple simple units, equation 2.1 can still be used — the difference is that the response is a vector of activities of multiple simple units, so the underlying probability distributions are multidimensional. Because we are not interested in the information about individual stimulus disparities, but rather how well all disparities are encoded, we integrate over the stimulus disparity,

$$I(\mathbf{R},\mathbf{S}) = \sum_{\delta}\sum_{i} p(\mathbf{r}_i|s_\delta)\log\frac{p(\mathbf{r}_i|s_\delta)}{p(\mathbf{r}_i)}. \tag{2.3}$$

We generated populations of simple units with (i) position shifts, (ii) phase shifts, or (iii) a combination of both (hybrid encoding). The Gaussian envelope width, σ, and the spatial frequency, f, were kept constant, and only the position x_0 and the phase ϕ parameters were allowed to vary.

We examined information encoded under two schemes. First, we computed the information under the assumption of uniformly spaced simple units. This ensures minimal overlap between the tuning curves of the simple units, and therefore avoids

redundancy (i.e. the suboptimal case where two or more units in the population have very similar tuning curves). Next, we examined information without imposing this uniform spacing, and allowed the simple units to assume random tuning profiles. We did this by generating 1,000 populations for which the position and/or phase shifts (according to the encoding mechanisms under evaluation) were randomly drawn from a uniform distribution. This yielded a distribution of information values for each of the mechanisms. As expected, we observed higher information values for the uniformly distributed population (Fig. 2.1D, horizontal lines) when compared to random populations (Fig. 2.1D, bar graph). In both cases, we found that hybrid populations carried the most information about the disparity imposed in our stimulus set (Fig. 2.1D).

Naturalistic binocular images

We generated naturalistic stereoscopic images using 100 light-field photographs extracted from the Light Field Saliency Database[197]. The dataset comprised images of a variety of indoor and outdoor scenes — representative stereo pairs are provided in Supplementary Figure 2.8 — and the corresponding depth maps. First, each RGB image (1080-by-1080 pixels) was converted to gray-scale values and down-sampled at the resolution of the corresponding depth map (328-by-328 pixels). Thereafter, we used the information provided by the depth map to render stereo pairs with arbitrary disparity range. From each light-field acquisition, we extracted a series of images focused at different points in depth, and rendered stereoscopic pairs by shifting the pixels of the original image by an amount proportional to the value of the depth map, restricting the maximum shift to 10 pixels. Pixels that were revealed behind occluded regions (by displacing image features in depth) were filled using linear interpolation. To prevent interpolation from affecting the training procedure, we excluded image patches for which more than 5% of the pixels were interpolated.

This method produced 200 stereo pairs. From these images we extracted 38,000 different pairs of smaller image patches (30-by-30 pixels). To ensure accurate disparity information, we excluded image patches with low variance of pixel intensity (gray level s.d. threshold = 20). All image patches were then scaled so that pixel intensity values were contained in the interval between –1 and 1, and randomly divided into

training and test sets, as described below.

We did not use standard two frame stereo datasets (e.g. Middlebury datasets) given that these contain a large range of disparities, making it difficult to obtain sufficiently large training sets for a given set of disparity values. We restricted the network to work on a small number of individual disparities for which we could provide training data. Rendering stereo pairs from the corresponding depth map, as described above, allowed us to generate images with arbitrary disparity range, and therefore increase the number of class exemplars available to train the network. Additionally, native two frame stereo datasets are typically composed of a comparatively small number of photographs, which could lead to exploring a narrow portion of the space of natural image statistics. This would affect the properties of the network and the degree to which it could generalize to other stimuli.

Binocular Neural Network (BNN)

Architecture

The binocular network was implemented using Theano[227], a library for efficient optimization and evaluation of mathematical expressions. We used a simple convolutional neural network that comprised (i) an input layer, (ii) a convolutional-pooling layer and (iii) an output logistic regression layer (Fig. 2.2A). The input is convolved with a series of kernels to produce one output map per kernel (which we refer to as convolutional maps). The use of convolution means that each kernel is applied at all different locations of the input space. This significantly reduces the number of parameters that need to be learned (i.e., we do not parametrize all possible pair-wise connections between layers) and allows the network to extract a given image feature at all different positions of the image.

Inputs were image patches (30x30x2 pixels; the last dimension carrying the left and right images) extracted from stereoscopic images. In the convolutional layer, binocular inputs are passed through 28 binocular kernels (19x19x2 pixels) producing 28 output maps (12x12 pixels). This resulted in 4,032 units (28 maps of dimensions 12x12 pixels) forming 2,911,104 connections to the input layer (4,032x19x19x2 pixels). Since this mapping is convolutional, this required that 20,244 parameters were learnt for this layer (28 filters of dimensions 19x19x2 plus 28 bias terms). We chose

units with rectified linear activation functions since a rectifying non-linearity is biologically plausible and necessary to model neurophysiological data[180]. The activity, a, of unit j in the k^{th} convolutional map was given by:

$$a_j^{(k)} = (w^{(k)} s_j + b_j^{(k)})_+ \tag{2.4}$$

where $w^{(k)}$ is the 19x19x2 dimensional binocular kernel of the k^{th} convolutional map, s_j is the 19x19x2 binocular image captured by the j^{th} unit, b_j is a bias term and $(.)_+$ denotes a linear rectification non-linearity (ReLU). Parameterizing the left and right images separately, the activity $a_j(k)$ can be alternatively written as:

$$a_j^{(k)} = (w^{(Lk)} s_j^L + w^{(Rk)} s_j^R + b_j^{(k)})_+ \tag{2.5}$$

where $w^{(Lk)}$ and $w^{(Rk)}$ represent the k^{th} kernels applied to left and right images (i.e. left and right receptive fields), while s_L^j and s_R^j represent the left and right input images captured by the receptive field of unit j.

The convolutional layer was followed by a max-pooling layer that down-sampled each kernel map by a factor of two, producing 28 maps of dimensions 6-by-6 pixels. Finally, a logistic regression layer (1,008 connections; 36 per feature map, resulting in 1,010 parameters including the bias terms) mapped the activities in the pooling layer to two output decision units. The vector of output activities r was obtained by mapping the vector of activities in the pooling layer a via the weight matrix W and adding the bias terms b, followed by a $softmax$ operation:

$$r = softmax(Wa + b) \tag{2.6}$$

The predicted class was determined as the unit with highest activity. For N-way classification, the architecture was identical except for the number of output units of the BNN.

Training procedure

The input stereo pairs were first randomly divided into training- (70%, 26,600 pairs), validation- (15%, 5,700 pairs) and test- (15%, 5,700 pairs) sets. No patches were simultaneously present in the training, validation and test sets. To optimize the BNN, only

the training and validation sets were used. We initialized the weights of the convolutional layer as Gabor filters with no differences between the left and right images. Therefore, initialization provided no disparity selectivity. With x and y indexing the coordinates in pixels with respect to the centre of each kernel, the left and right monocular kernels W^L and W^R of the j^{th} unit were initialized as

$$w_j^L = w_j^R = e^{-(x'^2+y'^2)/(2\sigma^2)} \cos(2\pi f x' + \phi) \tag{2.7}$$

with f=0.1 cycles/pixel, σ=3 pixel, θ=$\pi/2$ radians, $x' = x\cos(\theta) + y\sin(\theta)$, $y' = -x\sin(\theta) + y\cos(\theta)$, and ϕ the phase of the cosine term of each unit, which was equally spaced between 0 and π. The bias terms of these units were initialized to zero. During training we did not constrain the filters to any particular morphology, neither did we constrain properties such as spatial frequency selectivity. In the logistic regression layer, the weights and bias terms were all initialized to zero.

The BNN was trained using mini-batch gradient descent with each batch comprising 100 examples (50 examples of each class). For each batch, we computed the derivative of the loss function with respect to parameters of the network via back-propagation, and adjusted the parameters for the next iteration according to the update rule

$$w_{i+1} = w_i - \alpha \left\langle \frac{\partial L}{\partial w_{(D_i)}} \right\rangle \tag{2.8}$$

where α is the learning rate, and $\langle \partial L / \partial w_{(D_i)} \rangle$ is the average over the batch D_i of the derivative of the loss function with respect to the w, evaluated at w_i. The learning rate α was constant and equal to 0.001.

After evaluating all the batches once — completing one epoch — we tested the BNN using the validation image dataset. We repeated this process for a maximum of 1,000 epochs. Initially, the maximum number of iterations allowed without improvement was set to 10,000. To allow exhaustive optimization, this limit was increased by a factor of 2 every time there was an improvement of 0.5% in performance as tested in the validation set.

Evaluation

We tested the BNN using both natural and synthetic images. For natural images, we tested it using 5,700 held-out patches on the test image dataset (i.e. these exemplars were not used for training or validating the network). For comparison with neurophysiological observations, we also tested the BNN using random-dot stereogram patches. This test set consisted of 6,000 randomly generated stereograms containing a mixture of dark and bright dots on a gray background (dot size = 1 pixel; dot density = 50%).

For comparison with psychophysical observations, we also tested the BNN with large random-dot stereograms depicting a step-edge (240-by-240 pixels). The dot size was set to 8 pixels and the dot density was approximately 15%. No occlusion between the dots was allowed. The step disparity was set to 2 pixels. Disparity noise sampled from a Gaussian distribution (s.d.=8 pixels) was added to increase task difficulty. Stereograms could contain bright dots, dark dots (single polarity cases) or an even mixture of both (mixed polarity case) on a uniform mid-gray background. Bright, dark, and mid-gray pixels corresponded to values of $+1$, -1 and 0, respectively. Differences in the response to mixed and single polarity stereograms could be affected by differences in mean luminance or contrast. We sought to rule out such effects by performing control analyses where these properties were matched. In particular, we report the results obtained when the mean luminance (DC) was removed, as differences in DC can have a drastic effect on the population responses[200]. Similar results were obtained when single polarity stereograms were scaled to have the same peak-to-trough values (i.e. pixel intensities varied from -1 to $+1$, producing a range of 2), and scaled to match the range of the mixed polarity stereograms after we had removed the mean luminance. Supplementary Figure 2.11 compares results obtained with different manipulations of the images.

Modelling binocular receptive fields

The receptive fields of simple units in the BNN were not constrained to develop a particular structure (i.e. Gabor functions) during optimization — they could in principle develop any kind of morphology. We therefore assessed whether the receptive field structure mirrored that found in simple cells in primary visual cortex. In par-

ticular, we set out to test (i) if the receptive fields were well approximated by Gabor functions, and (ii) what kind of encoding mechanism they develop — i.e. position, phase or hybrid encoding.

We started by assessing whether the receptive fields were well approximated by Gabor functions. To reduce the number of free-parameters, we examined the horizontal cross-section of the receptive field, and fit a 1-dimensional Gabor function,

$$W = A \times e^{-(x-x_0)^2/(2\sigma^2)} \cos(2\pi f(x-x_0)+\phi). \tag{2.9}$$

We used a two-stage procedure for optimization. First, we ran a coarse grid-search to find a good initial guess for the parameters, whereby the combination of parameters with lowest sum of squared errors was selected. Then, taking the grid-search estimates as initial guesses, we estimated the final parameters using bound constrained minimization. The constrained parameters were the amplitude ($0 < A < +\infty$), the center of the envelope ($min(x) < x_0 < max(x)$), the phase ($-\pi < \phi < \pi$) and the frequency, which was constrained to an interval of $\pm 10\%$ around the peak of the Fourier transform of the receptive field profile. To assess whether disparity was encoded via position and/or phase shifts (Fig. 2.1B), we subtracted the position/phase parameters between the left and right receptive fields. The phase parameter was wrapped to $[-\pi, \pi]$.

To address consistency with neurophysiology, we examined the spatial frequency bandwidth of the receptive fields learnt by our model. We quantified spatial frequency bandwidth using two methods. First, we used a non-parametric approach of computing the spatial frequency tuning curve for each filter, and then determining the corresponding bandwidth (FWHM). We found that the spatial frequency bandwidth values were plausible when compared to the bandwidth of V1 neurons[228] (average bandwidth = 2.32 octaves; values ranged from 1.58 to 3.44 octaves). As a confirmatory procedure, we used a parametric approach based on the standard deviation and the frequency parameters of the Gabor fits. This yielded near-identical results, although 13/56 filters could not be evaluated using this method as they produced NaN estimates.

Varying the number of simple units and testing the importance of positional disparities

When defining the architecture of the BNN, we arbitrarily set the number of simple unit types to 28. To ensure that our results hold in a more generalized manner, we additionally trained similar versions of the Binocular Neural Network while varying the number of simple unit types. The remaining parameters of the network were kept constant. After optimization, we found a similar pattern of results: we achieved high classification accuracies (Supplementary Figure 2.9A), and the binocular receptive fields developed a combination of phase and position disparities (Supplementary Figure 2.9B, C).

Relating simple unit properties (i.e. their receptive fields) to the readout of their activity is a key step in understanding the computation performed by the network. We chose to deploy the network with 28 types of simple units as opposed to the models with fewer units. This was because it provided a richer substrate to determine the relationship between simple units properties and their readout, and allowed us to perform a 'lesion' analysis of the network where performance was not uniquely dependent on a very small number of units. With fewer units (e.g. 8), performance when dropping units would have become unstable.

Estimating correlated vs. anticorrelated amplitude ratios

Complex units in the BNN responded more vigorously to correlated (cRDS) than anticorrelated stereograms (aRDS) (Fig. 2.3A), a phenomenon that is observed in disparity selective V1 complex cells[87,89]. We examined whether the degree of attenuation observed in our network was compatible with electrophysiological data. Attenuation is commonly assessed by modelling tuning curves for aRDS and cRDS, and then evaluating the ratio between the corresponding amplitudes[87,121,214]. Therefore, we modelled the tuning curves using Gabor functions (similar to those used to model the binocular receptive fields) and computed the ratio between the amplitude parameter for correlated and anticorrelated stimuli. We started by generating disparity tuning curves for each complex unit by computing the activity elicited by correlated or anticorrelated random-dot stereograms (50% dot density) with disparities ranging from —20 to 20 pixels (100 trials per disparity) (Fig. 2.3B). To avoid relying on a single fit

per complex unit, we used bootstrapping to generate 5,000 resampled tuning curves, and we fit a Gabor to each sample. The average explained variance of the fits to the disparity tuning curves was $R^2 = 0.945$ ($R^2 = 0.93$ for cRDS and $R^2 = 0.96$ for aRDS). Based on these parameters, we computed the respective amplitude ratios by dividing the amplitudes for aRDS by the amplitudes for cRDS. We finally arrived at a distribution of amplitude ratios (Fig. 2.3C) by pooling the data across complex units.

N-way classification

In addition to the binary case, we also trained a network to perform N-way classification. The only change required to the network was an increase in the number of output complex units. In particular, we optimized a network for 7- and 11-way classification. In these cases, the complex units of the network also display inversion and attenuation for anticorrelated random-dot stereograms, with comparable but more variable amplitude ratios (Supplementary Figure 2.10). We found that the corresponding tuning curves featured abrupt changes in selectivity, and some were not well described by Gabor-like profiles. We note that this is also the case in cortex (i.e., that Gabor functions do not always describe disparity tuning well). However, the abrupt variations in tuning could be alleviated by varying the temperature of the *softmax* nonlinearity, or by defining the N-way classification problem to operate over a broader disparity space.

Computing optimal stimuli

To confirm that the model was well tuned to extract physical binocular disparities, we computed input images that could best activate the complex units of our model. The intuition is that we can visualize what inputs are most efficient in driving a given complex unit, and thereafter evaluate whether the input is sensible. The objective function is therefore the activity of a given complex unit, which we want to maximize. Equivalently, for an output unit j, we minimized the negative of its input:

$$L_j = -(W_j a + b_j) \tag{2.10}$$

where a is the vector of simple unit activities, W_j is the readout weight matrix for

the j^{th} complex unit, and b_j is the bias term. The goal is thus to find an input image that minimizes L_j (i.e. maximizes the complex unit activity; Fig. 2.4A). We did this via gradient descent: we started with a random noise input image, x, computed the gradient of the loss function with respect to the input image, and adjusted the latter according to the update rule:

$$x_{i+1} = x_i - \alpha \frac{\partial L}{\partial x} \tag{2.11}$$

where α is the step size (empirically set to 1). We limited the number of iterations to 100 as this was enough to ensure that optimization reached a stable image configuration (i.e. the correlation between the stimulus in to consecutive iterations saturated at 1).

The stimuli that best activated the complex units resembled contrast edges horizontally translated between the eyes, in the direction consistent with the preferred disparity of the complex unit (Fig. 2.4B). This is consistent with detecting positional offsets. The structure of the optimal stimuli was very similar across the eyes, indicating that stimuli with non-physical (i.e. phase) disparities are not ideal to activate the BNN's complex units.

Step-edge depth discrimination and depth-sign maps

In its original form, the BNN takes a 30-by-30 input image patch and produces a binary output corresponding to the predicted disparity (*near* or *far*). Once trained, however, convolutional neural networks can be applied to higher dimensional inputs, without requiring any changes in the parameters of convolutional layers. We took advantage of this convenience to test the BNN with larger binocular inputs. The only required modification to the BNN happened in the readout layer, where we applied the mean read-out weight for each simple unit in an element-wise manner. This resulted in two output activity maps — one for near disparities (*near* map), and another one for far disparities (*far* map). More formally, the vector of activities in the j^{th} output map was defined as:

$$a_{out}^{(j)} = \sum_{(k=1)}^{28} a_{conv}^{(k)} \hat{w}_{out}^{(kj)} + b^{(j)} \tag{2.12}$$

where $a_{conv}^{(k)}$ is the vector of activities in the k^{th} convolutional map, $\hat{w}_{out}^{(kj)}$ is the mean readout weight between the k^{th} convolutional map and the j^{th} output unit, and $b^{(j)}$ is the vector of bias terms of the j^{th} output unit. Finally, we combined the two output maps by element-wise subtracting the activities of the *near* map from the *far* map, so that positive values reflect higher *near* activity, while negative values reflect higher *far* activity.

Relationship between simple unit selectivity and readout

The activity of complex units in the network depends on the readout of the activity of the population of simple units. We assessed whether there was a relationship between the receptive fields of simple units and the corresponding readout weights. Take, for instance, the complex unit that responded to *near* stimuli: how does this complex unit combine the activity of the population of simple units? We found that it used readout weights that were proportional to the average interocular receptive field cross-correlation at *near* disparities (Supplementary Figure 2.13, red elements; Pearson's $R = 0.90$, $p < 10^{-9}$). In the same manner, the readout weights for the *far* complex unit were proportional to the average interocular receptive field cross-correlation at *far* disparities (Supplementary Figure 2.13, blue elements; Pearson's $R = 0.89$, $p < 10^{-9}$). The readout weight is therefore proportional to the interocular receptive field cross-correlation at the preferred disparity of the complex unit.

Derivation of the Binocular Likelihood Model

Interocular RF cross-correlation and disparity selectivity

It has been noted elsewhere that computing the cross-correlogram between the left and right receptive fields yields a very good approximation of the disparity tuning curve[66,71,187]. Below we present a derivation that describes this relationship. We start by considering the response r of binocular simple cells to a given binocular stimulus with disparity δ. The binocular half images (i.e., the images captured by the left and right eyes) are horizontally translated versions of one another. Thus, the stereo pairs presented in a given trial t can be defined as $\{S_t(x), S_t(x+\delta)\}$. As observed experimentally, the response of a binocular simple cell can be well described by linear

spatial filtering and rectification, followed by a non-linearity[63,70],

$$r = g([S_t(x)W_L(x) + S_t(x+\delta)W_R(x)]_+), \tag{2.13}$$

where $W_L(x)$ and $W_R(x)$ denote the receptive fields of the simple cell for the left and right eyes, and g is an expansive nonlinearity. It has been shown that this non-linearity is well described by a power law with an exponent of approximately 2, $g(x) = x^2$, for $x > 0$[70]. We assume an unrectified squaring non-linearity for mathematical convenience, however, similar results would be obtained for a rectifying squaring non-linearity[187]. Based on this, we can compute a disparity tuning curve, $f(\delta)$, by averaging the response of the simple cell across a large number of trials T,

$$\begin{aligned} f(\delta) &= \frac{1}{T}\sum_{t=1}^{T} r_t \\ &= \frac{1}{T}\sum_{t=1}^{T}\Big(S_t(x)W_L(x) + S_t(x+\delta)W_R(x)\Big)^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\Big((S_t(x)W_L(x))^2 + (S_t(x+\delta)W_R(x))^2 \\ &\quad + 2S_t(x)W_L(x)S_t(x+\delta)W_R(x)\Big). \end{aligned} \tag{2.14}$$

As many others have noted[70,186,189,191], the first two terms are monocular and do not depend on binocular disparity – over many trials, these two terms should be a positive constant, C, independent of the disparity δ of the stimulus. The disparity dependent modulation of the tuning curve is captured by the interaction term,

$$f(\delta) = \frac{1}{T}\sum_{t=1}^{T} 2S_t(x)W_L(x)S_t(x+\delta)W_R(x) + C. \tag{2.15}$$

This expression describes the expected response for a simple cell with receptive fields $W_L(x)$ and $W_R(x)$ to stereoscopic pairs that are translated horizontally in relation to one another by a given disparity δ. Under this formulation, the response of the simple cell is proportional to the stimulus unnormalized cross-correlation,

$S_t(x)S_t(x+\delta)$, weighted by the product of the left and right receptive fields, $W_L(x)W_R(x)$, known as the binocular interaction field[70].

However, as we will now show, it is useful to reformulate this expression. Because the stereoscopic pairs are simply translated in relation to the position of the receptive fields, it is equivalent to compute a disparity tuning curve by applying the horizontal shift to the receptive fields, while keeping the stereoscopic images in the same horizontal position ($a(x-\delta)b(x) = a(x)b(x+\delta)$),

$$f(\delta) = \frac{1}{T}\sum_{t=1}^{T} 2S_t(x)W_L(x)S_t(x)W_R(x-\delta) + C \tag{2.16}$$

$$= \frac{1}{T}\sum_{t=1}^{T} 2S_t(x)^2\ W_L(x)W_R(x-\delta) + C. \tag{2.17}$$

Equation 2.17 is convenient because it expresses the disparity tuning curve as a function of the dot product between the left and right receptive fields, translated according to the disparity δ. This is by definition the cross-correlation between the left and right receptive fields $(W_L \star W_R)[\delta]$. Note that $\frac{1}{T}\sum_{t=1}^{T} S_t(x)^2$ is simply the average energy of the stimulus over T trials, which influences the amplitude of the tuning curve (but not its morphology). Therefore,

$$f(\delta) = 2(W_L \star W_R)[\delta]\ \frac{1}{T}\Big(\sum_{t=1}^{T} S_t(x)^2\Big) + C \tag{2.18}$$

$$= 2(W_L \star W_R)[\delta]\ \mathbb{E}\Big(S_t(x)^2\Big) + C. \tag{2.19}$$

This formulation provides a mathematically convenient way of expressing tuning for binocular disparity solely based on the receptive fields of simple units. Next, we will take advantage of this convenience to establish a relationship between simple unit properties and their readout by complex units.

Optimal readout of simple unit activity by disparity selective complex units

In the previous section, we showed that the disparity tuning curve of a simple unit can be well approximated by the scaled cross-correlogram between the left and right

receptive fields. We also suggested that stimulus contrast energy induces variability in the firing rate of simple units. This high variability makes simple units unsuitable for the detection of depth. By combining the activities of multiple simple units, complex units provide much better estimates of disparity. The classical disparity energy model obviates this problem by combining the outputs of four simple units with the same preferred binocular disparity, but with their receptive field phase in quadrature[63].

We now ask how could we optimally combine the activities of a population of simple units with highly variable firing rates. Here, we consider not only the variability in firing rate statistics, but also extrinsic variability induced by the stimulus. Inspired by previous work on optimal sensory representations[229], we tackle this problem from a probabilistic viewpoint. Let us interpret the distribution of activity of a simple cell i given a particular disparity δ as describing the likelihood of observing the firing rate r_i given the disparity δ. We make the simplifying assumption that the response of a simple unit, affected by intrinsic and extrinsic variability, follows a Gaussian distribution around the mean firing rate value, which is given by the corresponding tuning curve, $f_i(\delta)$. Thus, the likelihood for a given simple cell i is given by

$$p(r_i|\delta) = \frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{(r_i - f_i(\delta))^2}{2\sigma_i^2}}. \tag{2.20}$$

This equation expresses the probability of observing a firing rate r_i given a stimulus with disparity δ. Assuming independence across a population of N simple cells, we can now combine these probabilities to obtain a joint likelihood,

$$L(\delta) = p(\mathbf{r}|\delta) = \prod_{i=1}^{N} p(r_i|\delta). \tag{2.21}$$

By working in log-space, we can convert the logarithm of the product of likelihoods into a sum of logarithms of the likelihood. This is useful because we can express the computation of the likelihood as sum over the activity of many neurons, which is a biologically plausible operation. Equation 2.21 thus becomes

$$\log L(\delta) = \sum_{i=1}^{N} \log p(r_i|\delta) \tag{2.22}$$

$$= \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{(r_i - f_i(\delta))^2}{2\sigma_i^2}}\right) \tag{2.23}$$

$$= \sum_{i=1}^{N} -\frac{(r_i - f_i(\delta))^2}{2\sigma_i^2} - \log(\sqrt{2\pi\sigma_i}) \tag{2.24}$$

$$= \sum_{i=1}^{N} \frac{r_i f_i(\delta)}{\sigma_i^2} - \frac{1}{2}\left(\frac{r_i^2}{\sigma_i^2} - \frac{f_i(\delta)^2}{\sigma_i^2} - \log(2\pi\sigma_i)\right). \tag{2.25}$$

$$\tag{2.26}$$

The second term in equation 2.26 can be ignored if we assume that the tuning curves of the population of simple cells cover homogeneously the disparities of interest, and thus $\sum_{i=1}^{N} f_i(\delta)^2 = constant$. Therefore, dropping the quantities that do not depend on the disparity δ, the computation of the log-likelihood simplifies to a sum of the products between the observed simple cell firing rates r_i, and the corresponding tuning curves, $f_i(\delta)$,

$$\log L(\delta) = \sum_{i=1}^{N} r_i f_i(\delta). \tag{2.27}$$

While this is a useful formulation (and technically more generalizable), it is more intuitive to relate readout to binocular correlation. As we observered earlier, the cross-correlogram is a good approximation to the disparity tuning curve of individual simple cells. By replacing $f_i(\delta)$ according to equation 2.19 and dropping the constant term that does not depend on disparity, the log-likelihood can be written as

$$\log L(\delta) = \sum_{i=1}^{N} r_i \, (W_L \star W_R)_i[\delta]. \tag{2.28}$$

Therefore, a population of complex cells can approximate the log-likelihood over disparity simply by weighting the firing rates of individual simple cells by their in-

terocular receptive field cross-correlation. While this particular solution is specific to the assumption of Gaussian variability, the approach followed here could be applied to other forms of response variability using a suitably transformed version of the cross-correlogram. If one assumes Poisson variability, so as to model intrinsic firing rate variability, then the readout form would be a log-transform of the interocular receptive field cross-correlation.

It should be noted that this derivation approximates the behaviour of the BNN because Equation 2.14 used a squaring non-linearity while the BNN used a linear rectification. While this would produce differences in activity, the fundamental response properties are likely to be preserved between this derivation and the BNN.

Finally, we provide an example of a disparity tuning curve obtained using this simple analytical expression. In this simulation, we used 9 simple unit maps with Gabor receptive fields (f=0.0625 cycles/pixel; spatial frequency bandwidth, b=1.5 octaves; σ=6.27 pixels), covering the full combination of three position disparities ($\Delta x_0 = \{-3, 0, 3\}$ pixels) and three phase disparities ($\Delta\phi = \{-\pi, -\pi/3, \pi/3\}$ radians). Apart from the number of simple units, we kept the architecture of the model consistent with the Binocular Neural Network. Therefore, the output layer consisted of two complex units – one preferring *near*, the other preferring *far* disparities. The readout weights between simple and complex units were defined according to the analytical expression for our model (Equation 2.28). This instantiation of the model produced complex units with disparity tuning curves that closely resemble those of complex cells in V1 (Supplementary Figure 2.14A): the tuning curves for correlated and anticorrelated stereograms are well approximated by Gabor functions, and anticorrelated tuning curves are inverted and attenuated in relation to correlated stereograms.

The simple units in this instantiation of the model shared the same spatial frequency preference. This demonstrates that our model does not rely on spatial frequency pooling to produce attenuation in response to aRDS. The spatial frequency bandwidth of the output complex unit was smaller than that of the corresponding simple units (1.07 octaves), consistent with the findings that pooling activity across space narrows spatial frequency selectivity[79]. However, our model could also encompass simple units with multiple spatial frequencies, and their activities could be subsequently readout by complex units using the relationship established in equation

2.28. In this case, pooling across multiple spatial frequencies would increase the bandwidth of the output complex units, while further reducing the response of the model to spurious disparities[191] and sharpening the degree of disparity selectivity[78,79].

One prediction stemming from our model is that response saturation in *simple cells* could modulate the amplitude ratio of downstream complex cells. In particular, introducing a compressive nonlinearity at the level of *simple cells* — for instance, to account for sublinear binocular integration[230] — causes the aRDS response to further attenuate relatively to the response to cRDS. We demonstrate this effect in Supplementary Figure 2.14B. An expansive non-linearity at the level of simple cells, on the contrary, would cause the degree of attenuation to decrease.

Quantification and Statistical Analysis

We used bootstrap resampling and we report the corresponding 95% confidence intervals unless otherwise noted. Results were pooled across stimuli or units within a model, but not across different instantiations of models. For the results of fitting procedures, we report the proportion of variance explained by the models.

Data and Software Availability

We performed all analyses in Python (`http://python.org`) using standard packages for numeric and scientific computing. The data used for model optimization and implementations of the optimization procedure are available at `https://doi.org/10.17863/CAM.8538`.

2.5 Supplementary Figures

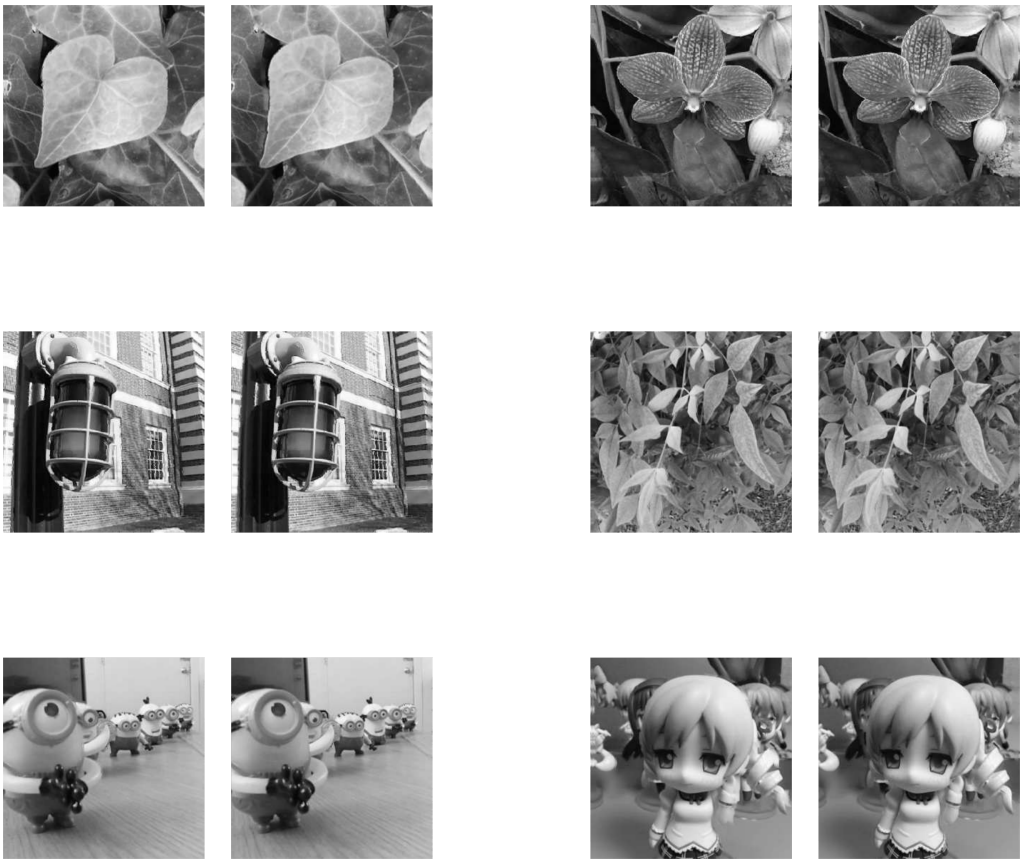


Figure 2.8: Examples of images used to train the binocular neural network (BNN). Related to Figure 2.2. Images were extracted from the Light Field Saliency Database [197], available at `http://www.eecis.udel.edu/~nianyi/LFSD.htm`. Stereo pairs are rendered for cross fusion.

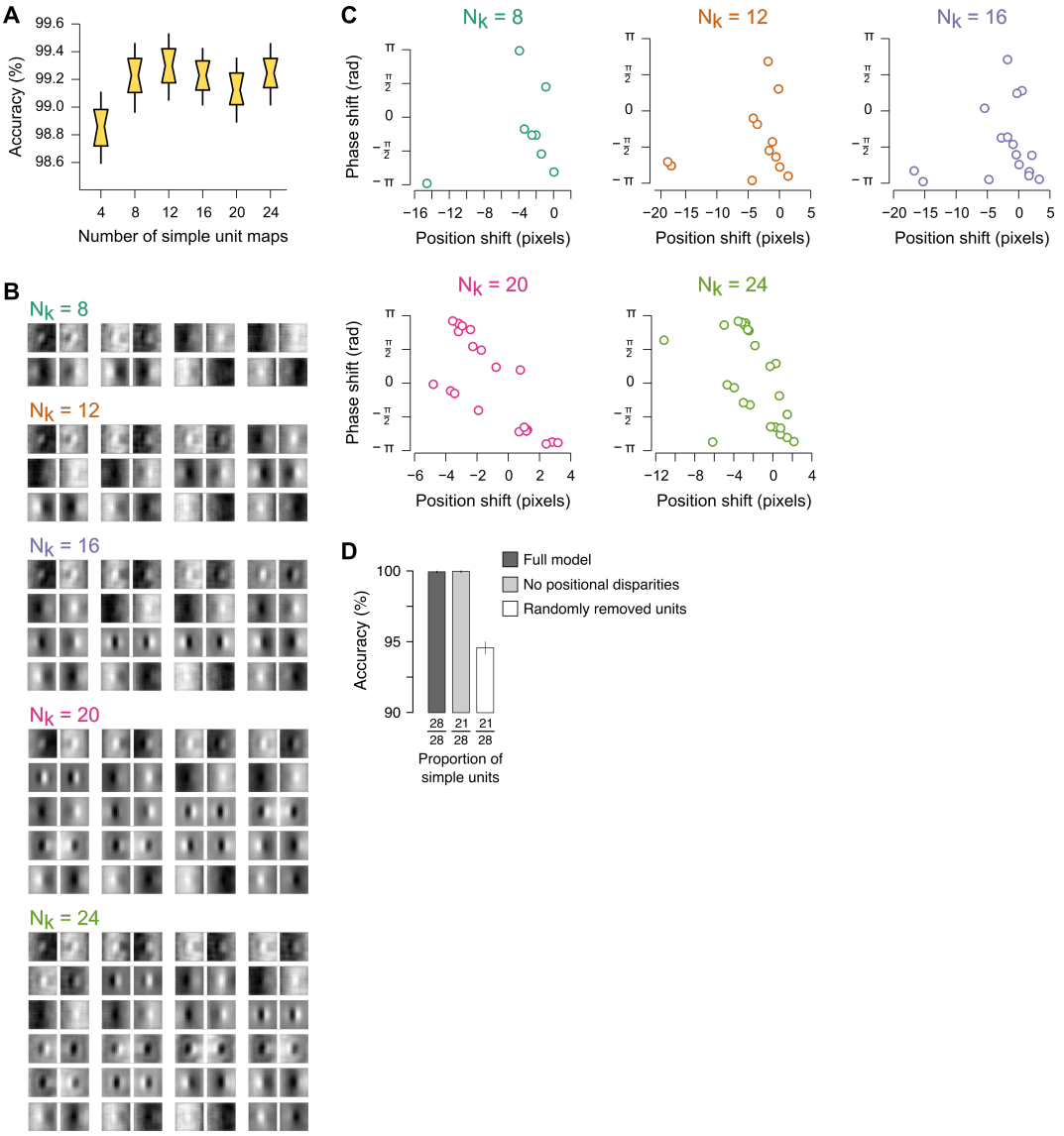


Figure 2.9: Varying the number of simple units in the network. Related to Figure 2.2. **(A)** Decoding accuracy for instantiations of the network with different number of simple units. **(B, C)** Binocular receptive fields developed by the corresponding instantiations, and the respective position and phase disparities. **(D)** Testing the importance of the positional disparity units for the 28 simple unit BNN. Decoding performance is shown for the full model, the model with the positional units removed (i.e., 25% of units around zero phase offsets), and the model with a randomly selected removal of 25% of the non-positional simple units (mean of 1,000 resamples). The limited impact of removing the position disparity units suggest these units do not play a strong role in determining the performance of the BNN.

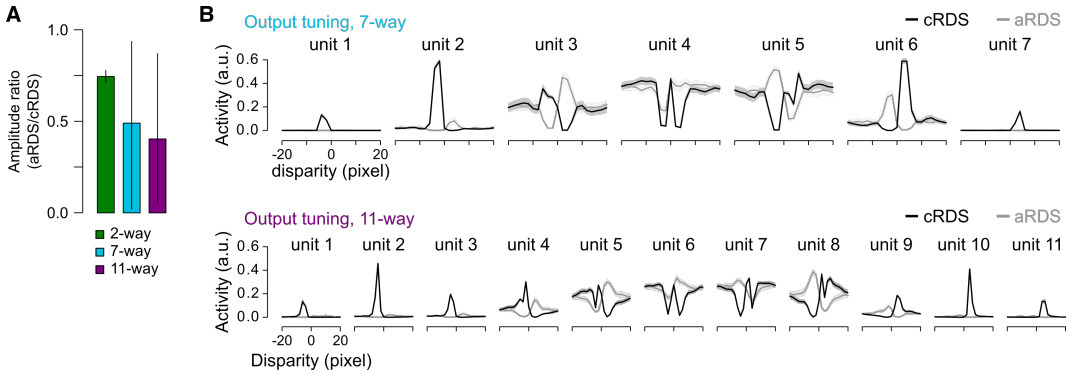


Figure 2.10: Comparing responses to correlated and anticorrelated stereograms. Related to Figure 2.3. **(A)** Response attenuation for anticorrelated versus correlated random-dot stereograms for 7- and 11-way classification. Bar graphs depict amplitude ratio (aRDS/cRDS) calculated based on peak-to-peak differences for 2-way, 7-way and 11-way (mean and $CI_{68\%}$ obtained via bootstrapping, 5,000 resamples per output unit). A bias towards values below unity is evident, consistent with the results shown in Figure 2.3C. Note that attenuation appears greater for 7- and 11- way classification. This may result from the network developing more sharply tuned units. Figure 2.3C indicates a hitherto unappreciated difference between the electrophysiological recordings of Cumming & Parker[87] vs. Samonds et al[89] at low attenuation ratios. Cumming & Parker[87] found a bias towards low-amplitude ratios. We speculate that this arose because they sampled closer to the fovea (RFs had eccentricities ranging from 1 to 4 degrees whereas Samonds et al sampled at 4 degrees eccentricity). This difference in sampling strategy may have resulted in recording neurons that were more sharply tuned for disparity, and thus showed greater attenuation for anticorrelated stimuli. **(B)** Disparity tuning curves of output units for 7-way (top) and 11-way classification (bottom). Inversion and attenuation can be observed in the majority of the units. There is considerable variability in the degree of attenuation across units, consistent with neurophysiological observations.

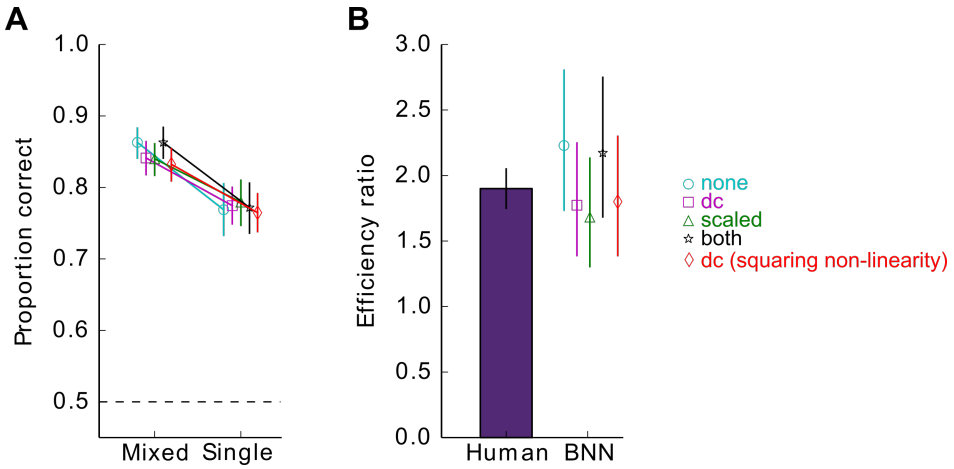


Figure 2.11: Control analyses for performance on mixed *vs.* single polarity stereograms. Related to Figure 2.5. We tested the performance of the BNN under several image adjustment conditions: no image adjustment (none); DC correction, in which we subtracted the mean intensity of the input images (dc); scale matching, in which we scaled the images to have the same peak-to-trough range (scaled); and DC correction plus scale matching, in which we removed the mean intensity of the images and scale them to the same peak-to-trough range (both). In the main paper we report results obtained with DC correction, and we re-plot them here to facilitate comparison with the remaining conditions. We ran an additional control in which the rectified-linear nonlinearity was replaced by a (unrectified) squaring nonlinearity **(A)** Proportion of correct trials under the different image adjustment conditions (1,000 trials; error bars show bootstrapped $CI_{95\%}$, 5,000 resamples). **(B)** Efficiency ratios for the BNN (1,000 trials; error bars depict $CI_{95\%}$, 5,000 resamples) plotted alongside experimentally measured efficiency ratios in humans (mean $\pm$ 1 s.d.)[199].

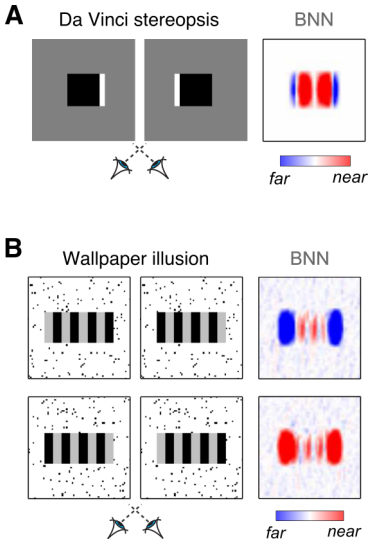


Figure 2.12: Response of the binocular neural network to half-occluded and wallpaper stimuli. Related to Figure 2.6. **(A)** Response of the BNN to unpaired bright flanks around a dark occluder. Luminance configuration was inverted relatively to Figure 2.6B. **(B)** Response of the BNN to a wallpaper pattern in which stripes are darker than the background. Compare to Figure 2.6C, where the pattern is defined by bright and dark stripes on a mid-gray background.

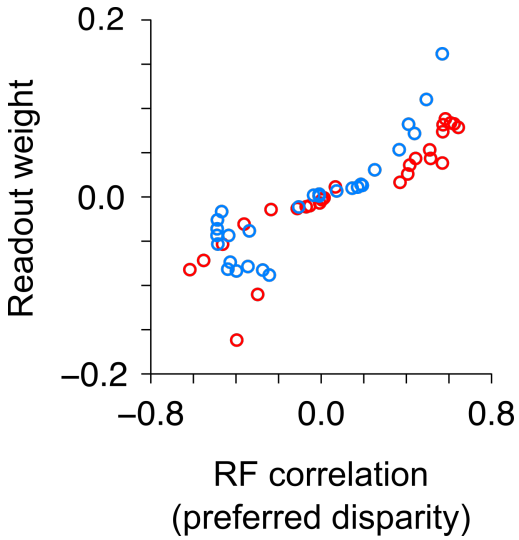


Figure 2.13: Relationship between the simple unit receptive fields and readout weights. Related to Figure 2.7. Examining the properties of the BNN showed that the simple unit readout weights are proportional to the receptive field interocular correlation at the preferred disparity. Thus, a complex unit with preferred disparity δ reads the activity of a simple unit with a weight proportional to the correlation between the simple unit's left and right receptive fields at the disparity δ. Red elements: near complex unit; blue elements: far complex unit.

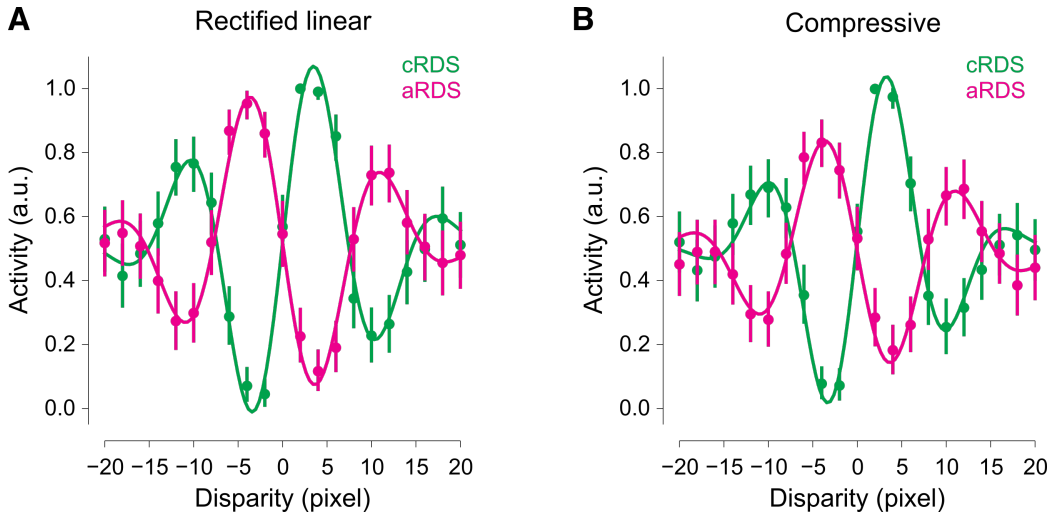


Figure 2.14: Disparity tuning curves obtained in a simple instantiation of the Binocular Likelihood Model (BLM). Related to Figure 2.7. Tuning curves were computed for correlated (green elements) and anticorrelated (pink elements) stereograms. Solid lines represent Gabor fits. Error bars depict resampled $CI_{95\%}$ (5,000 resamples). **(A)** Tuning curves for cRDS and aRDS assuming simple units with linear rectification **(B)** As in **A**, but assuming simple units with a compressive non-linear activation function (in this case, the square root) - i.e. effectively implementing sublinear binocular integration [230]

Chapter 3

What you don't see can hurt you: how imperceptible signals shape what we see

3.1 Introduction

Sir Charles Wheatstone[12] aquatinted the world with the glories of binocular vision. Through practical demonstration of the stereoscope, he broadcast his insight that differences between the two eyes' retinal images provide precise information about three-dimensional (3D) shape. Modern treatments of stereopsis, in both artificial- and biological- systems, have emphasised the need to identify similarities between the eyes to understand 3D vision. This is captured as the "binocular correspondence problem"[16,158,194] that entails matching the same image features between the two eyes so that the difference in retinal positions can be extracted.

The potential difficulties of binocular matching were made stark by random dot stereogram (RDS) displays[16]. Here, observers perceive a coherent 3D shape, despite viewing monocular images that contain no meaningful structure. Given that the elements making up the display are self-similar, there is huge potential for the brain to match up the "wrong" parts of the display. Yet, viewers typically perceive compelling 3D forms without much difficulty. The logical interpretation of these stimuli was, therefore, that the brain is faced with a hard problem in identifying the one "correct

match" in a sea of "false matches" that conflate signals originating from locations in 3D space.

This intuition has guided thinking to ask how the brain could find the correct match. However, the more pertinent question is to consider the best way for neurons to extract statistical information about the depth structure of the scene[231]. In particular, potentially valuable information is conveyed by neural signals that appear discordant with perceptual experience. Moreover, the properties of binocular neurons are not particularly well-suited to identifying "correct" matches per se: many respond best to different features in the two eyes[67,71,73]. This includes tuning to "anticorrelated" stimuli (Fig. 3.1a) that are contrast-inverted between the eyes (i.e., a bright feature in one eye's image corresponds to a dark feature in the other)[63,87].

On the basis that the brain seeks to identify "correct" matches, it is very puzzling that the brain responds to anticorrelated stimuli: in general, a single physical feature in the environment will not project opposite contrast images into the two eyes. Moreover, viewing anticorrelated stimuli leads to a discombobulating perceptual experience such that it has long been thought that observers are essentially blind to the information depicted in anticorrelated RDS[33,34,37,38,150].

Here we demonstrate that the traditional focus on "correct" matches fails to capture the signals used by human visual system. In particular, we show that anticorrelated signals — traditionally viewed as "false" — are, in fact, perceptually informative. We do this by mixing correlated and anticorrelated elements together to reveal profound consequences for perception. Specifically, we provide evidence that the disparity information carried by anticorrelated features (itself imperceptible) masks correlated features. This masking effect is contingent upon (i) the disparity of the anticorrelated features, (ii) the magnitude of the disparity, (iii) the spatial configuration of binocular correlation, and (iv) the relative onset timing of anticorrelated and correlated elements.

This demonstrates the utility of signals that we know are carried by V1 neurons, but which, to this point, have been regarded as a nuisance signal that should be filtered out. We show that the brain is sensitive to information that we "cannot see" and it uses these signals to shape perceptual judgments. This is compatible with the idea that neural sensors for mismatched features provide a principled means of acquiring more information about the depth structure of the viewed scene[231]. These

mismatches provide proscriptive "what not" information to drive suppression of unlikely interpretations of the viewed scene.

3.2 Methods

Participants

Participants had normal or corrected-to-normal vision, were screened for stereo deficits, and provided written informed consent. All procedures were approved by the University of Cambridge ethics committee. For Experiment 1, we tested 12 participants (4 females; aged between 21 and 41 years). Three participants (including the two authors) were tested in experiments 2-4.

Main experiment

In Experiment 1, we presented stereoscopic stimuli using a stereoscope equipped with two Samsung 2233 LCD displays (1680 x 1080 pixels, refreshed at 120 Hz) driven by an NVIDIA Quadro 4000 graphics card. The monitors were placed laterally and were viewed using a pair of infrared permeable mirrors. The viewing distance was 50 cm, yielding a resolution of 31 pixels per degree of visual angle. The position of each eye was recorded using an EyeLink 2000 video eye tracker at a sampling rate of 1000 Hz.

We used MATLAB (The Mathworks Inc., Natick, USA) and Psychtoolbox[232–234] for generating and delivering stimuli. Stimuli were random-dot stereograms (RDS) composed of black and white dots on a mid gray background. The RDS (7°x7°) was presented within a circular aperture (11° radius) surrounded by a pink noise pattern. Each dot subtended approximately 4 minutes of arc and the dot density was 96 dots/deg^2. No occlusion between the dots was permitted. Stimuli were presented for 300 milliseconds. The displays were luminance calibrated and the display outputs linearised.

The dots in the RDS could be either correlated or anticorrelated between the eyes. Binocular disparity was defined according to a step edge configuration, where the left and right halves of the stereograms had opposite disparity sign (Fig. 3.1a; 3 arcmin). We rendered stimuli such that (i) correlated and anticorrelated dots had the same disparity configuration (Fig. 3.1a, "same" condition) or (ii) correlated and

anticorrelated dots had the opposite disparity configuration (Fig. 3.1a, "opposite" condition). Observers were asked to report which half of the stereogram they perceived closest in depth. We varied the proportion of correlated to anticorrelated dots according to a QUEST procedure[235] set to estimate 75% correct thresholds for the "same" and "opposite" disparity sign conditions (2 runs; 100 trials per run). To control for the possibility that anticorrelated dots carried a perceivable depth signal per se, we collected 200 additional trials where the QUEST stimuli were "replayed" exclusively with the correlated dots or the anticorrelated dots. We then quantified the proportion of correct responses for correlated and anticorrelated stimuli.

Subsequent experiments

Experiments 2-4 sought to characterize the masking effect with high spatial and temporal precision. To this end, we used a modified Wheatstone stereoscope with a long viewing distance (170 cm) and high temporal resolution display (ViewPixx 3D monitor, VPixx Technologies Inc., Saint-Bruno, Canada) driven by a NVIDIA Quadro FX 5600 (1920 x 1080 pixels; refresh rate = 120 Hz). The left and right images were presented side-by-side on a single screen, ensuring that the binocular images were synchronized. The display resolution was 110 pixels per degree.

RDS stimuli (5°x5°) were presented within a central aperture (8°) surrounded by a random pink noise pattern. Each stereogram contained a mixture of black and white dots on a mid gray background. Each dot subtended approximately 4 arcmin and the dot density was 85 dots/deg^2. No occlusion between the dots was permitted. On each trial, stimuli were presented for 300 milliseconds. The display was luminance calibrated and the display outputs linearised.

Experiment 2: The effect of disparity offset between correlation and anticorrelation in depth

In Experiment 1, we contrasted performance when correlation and anticorrelation carried the same or opposite disparity signs. In Experiment 2, we parametrically varied the difference between the disparities of correlated and anticorrelated dots. The correlated dots depicted a step-edge configuration and their disparity was kept constant (±3 arcmin). We varied the disparity given to anticorrelated dots parametrically

(13 disparities equally spaced between ±9 arcmin), such that the difference between correlated and anticorrelated disparities ranged from 0 to 12 arcmin.

To examine if the effect depended on disparity magnitude, we ran additional experiments where we increased the magnitude of the correlated disparity (1.5, 9 and 15 arcmin). As before, we varied the disparity of the anticorrelated dots in a graded manner. The distance between correlated and anticorrelated disparities ranged from 0 to 6 arcmin for the fine magnitude (1.5 arcmin), from 0 to 18 arcmin for the intermediate magnitude (9 arcmin) and from 0 to 30 arcmin for the large magnitude.

The proportion of correlated dots in the stimulus was individually adjusted for each observer. We chose the proportion of correlated dots such that the variation in anticorrelated disparity had a clear effect in performance. Therefore, we defined the proportion of correlated dots to be the mean of the 75% correct thresholds for when correlated and anticorrelated dots have the same disparity sign ("same" condition in Experiment 1) and opposite disparity signs ("opposite" condition in Experiment 1). Thresholds were estimated using a QUEST adaptive staircase, and each observer performed 4 runs (260 trials) per disparity magnitude.

Experiment 3: Spatial interaction between correlation and anticorrelation

We sought to characterize the spatial extent (in the frontoparallel plane) over which correlated and anticorrelated disparities could interact. We manipulated the spatial periodicity between correlation and anticorrelation according to a horizontally oriented square-wave grating (Fig. 3.5a). The square-wave grating could have one of 7 spatial periods equally spaced in logarithmic space, ranging from 1 to 70 arcmin. As in Experiment 1, observers were asked to report the configuration of the step-edge, and anticorrelated dots could have the same or opposite disparity sign relative to correlated dots (Fig. 3.1a, "same" and "opposite" conditions; disparity = 3 arcmin). Finally, we calculated the proportion of correct responses for these two conditions as a function of spatial period.

Experiment 4: The effect of onset asynchrony between correlation and anticorrelation

We sought to examine the effect of temporal onset asynchrony between correlation and anticorrelation on the performance of the observers. We varied the temporal onset of anticorrelated and correlated dots, such that anticorrelated dots could precede or follow correlated dots (Fig. 3.6a). We tested onset asynchronies between 8 and 33 milliseconds (1 and 4 refresh frames, respectively).

Three observers participated in the experiment. Stimuli were RDS depicting a step-edge in depth and the task was to report the configuration of the step-edge in depth (disparity $= \pm 3$ arcmin). On each trial, correlated and anticorrelated dots were presented for 133 milliseconds, regardless of the onset asynchrony (we also obtained results for one with a presentation time of 250 milliseconds). The proportion of correlated dots was adjusted so that observers were 75% accurate without temporal asynchrony (QUEST procedure, as in Experiment 1). To rule out a general masking effect, each observer performed an additional condition in the anticorrelated dots were replaced by uncorrelated dots (i.e. unpaired dots, for which disparity is not defined). We analysed the proportion of correct responses as a function of onset asynchrony.

Data analyses

Analyses were performed in MATLAB (The Mathworks Inc., Natick, USA). For experiment 2, psychometric functions were fit using psignifit 3.0[236].

Eye tracking

During the main experiment, binocular eye position was measured using an Eyelink 2000 eye tracker (SR research) that imaged the eyes through a pair of infrared permeable mirrors. Prior to each experimental block, we ran a calibration procedure where participants fixated in a set of points with known coordinates in the visual field. We used general linear modelling to fit these reference coordinates to the raw eye position measurements in camera coordinates. The resulting bias and scale parameters were then used to calibrate the data collected in the experimental block that followed. The data was then segmented into trials beginning 100 ms prior to stimulus

onset and ending 400 ms after stimulus offset. We used the data from the Eyelink's in-built blink detection to reject trials over which blinks had occur (±100 ms around the blink onset and offset). Thereafter, we used a k-nearest neighbours based outlier detection algorithm to exclude outlier trials (on average 5.5% of the trials). We quantified changes in vergence by comparing the vergence measurements at each time point within the stimulus presentation period with the mean vergence measurement in the 100 milliseconds preceding stimulus presentation.

Code availability

Code to reproduce the results reported in this article is freely available at `https://github.com/nrgoncalves/antidisparity`.

3.3 Results

Our starting hypothesis was that the brain exploits anticorrelated signals to rule out specific interpretations of the viewed stimulus. To test this idea, human observers performed a depth discrimination task ("which side of the image is closer to you?") when viewing stereograms containing a mixture of correlated and anticorrelated features. We began by testing two extreme situations: either correlated and anticorrelated dots indicated the same depth; or they specified opposite configurations. If anticorrelation is not used by the visual system, as most studies suggest[237], then observers should not be affected by the signals depicted by anticorrelation. In particular, while adding anticorrelated signals to a correlated display would degrade overall performance, its effects would not be specific to the disparity of the anticorrelated features. By contrast, if anticorrelation is used to rule out specific interpretations, we expect poorer perceptual performance when the disparities of correlated and anticorrelated dots coincide. We find that the disparity of the anticorrelated features has a highly specific effect on stimulus visibility.

We demonstrate the logic of our paradigm in Figure 3.1. We first render a depth edge using exclusively correlated dots (Fig. 3.1b): the reader should easily perceive a step in depth (the left side of the stimulus should appear closer to the reader when wearing red-cyan anaglyph glasses with the red filter over the left eye). We next gen-

erate two stereograms using exclusively anticorrelated dots (Fig. 3.1c,d): in this case, the reader should experience a confusing mess with no clear depth percept. Note that the stimuli in 3.1c and 3.1d are perceptually indistinguishable despite containing very different disparities. Finally, we mix the correlated and anticorrelated dots together (Fig. 3.1e,f). This makes the overall depth harder to see (i.e., in comparison with Fig. 3.1b), however visibility depends on the disparity conveyed by anticorrelated features. In the case where correlated and anticorrelated dots depict the same disparity (Fig. 3.1e), a striking masking effect occurs and the observer should struggle to perceive depth in this stimulus. However, the masking effect is weaker when the correlated and anticorrelated disparities depict different depths (Fig. 3.1f). Importantly, the only difference between Figs. 3.1e and 3.1f is the disparity of the anticorrelated dots, which — as demonstrated in Figs. 3.1c and 3.1d — are perceptually indistinguishable. This indicates that the disparity information carried by the anticorrelated stimuli (traditionally understood as 'false matches' that do not inform perception) is critically important in determining perceptual visibility.

We quantified this effect in twelve observers by measuring discrimination thresholds and contrasted performance in 'same' and 'opposite' conditions. We found that every observers' performance was worse when correlated and anticorrelated dots in the stimulus depicted the same disparity (Fig. 3.2a; two-tailed paired t-test, $t(11)=6.00$, $p<0.001$; difference 95% CI=[0.07, 0.15]).

To ensure that observers had no residual sensitivity to the disparity information carried by anticorrelation, we used a control condition in which only the anticorrelated components of the stereograms viewed in the main experiment were presented. In this case, observers performed at chance (two-tailed one sample t-test against chance performance, $t(11)=0.30$, $p=0.77$; 95% CI=[0.47, 0.54]), indicating that they could not extract depth based on the anticorrelated signals alone (Fig. 3.2b). Further, we recorded eye movements and found no evidence for a differential effect of the anticorrelated disparity signals on binocular eye position (Fig. S1), making a non-specific explanation for our findings unlikely.

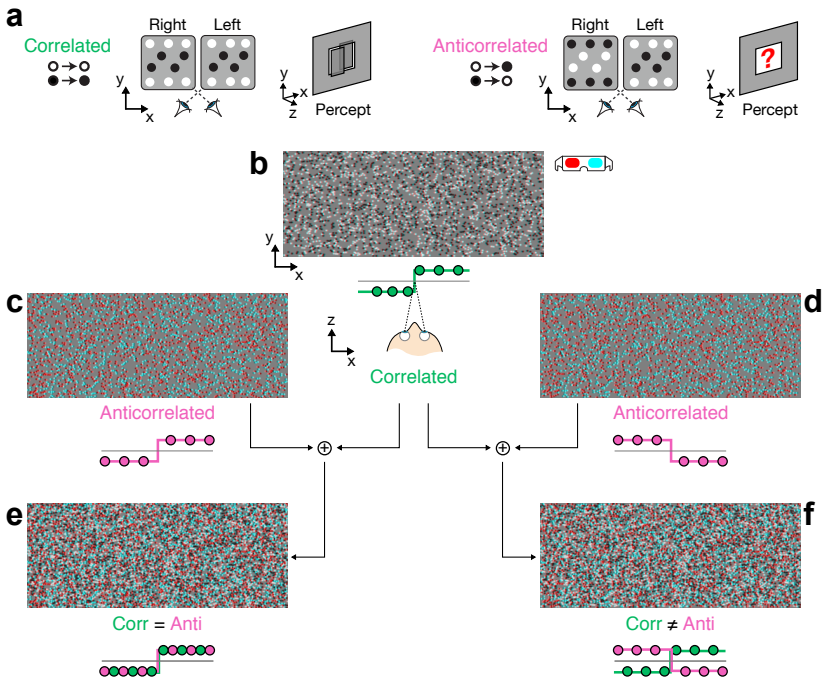


Figure 3.1: Perception is affected by disparity of anticorrelated features. (a) Cartoon illustrations of pure correlated vs. anticorrelated stereograms. In the correlated case, dark dots in one eye's image match dark dots in the other, and bright dots match bright dots. When the images are binocularly combined (here stereopairs shown for cross-eyed fusion), a percept of a step edge in depth emerges. In the anticorrelated case, bright dots in one eye are paired with dark dots in the other; this does not support a clear impression of depth. (b) A step-edge depicted using a correlated random dot stereogram (RDS) for viewing through red-cyan anaglyphs (red filter over left eye). The icon underneath the RDS depicts the spatial arrangement of the disparity signals. (c, d) Anticorrelated RDS depicting the same (c) and opposite (d) disparity configurations as in panel b. No impression of depth is elicited from these stimuli, and they are perceptually indistinguishable from each other, despite containing different disparity arrangements. (e, f) RDS mix the correlated and anticorrelated stereograms with the same (b + c) or opposite (b + d) disparity configurations. While it is harder to see the depth than in (b), the vast majority of observers report that the depth percept is clearer in (f) than in (e).

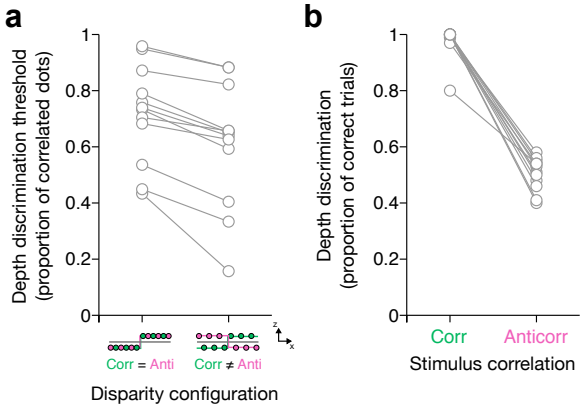


Figure 3.2: Masking is stronger when correlated and anticorrelated features depict the same disparity. Participants (N=12) judged whether the step was closer on the left or right side of the viewed display. We used an adaptive staircase procedure to measure discrimination thresholds for which we varied the proportion of anticorrelated dots in the display. (a) Proportion of correlated dots tolerated by the observers (75% correct thresholds) for stimuli where correlated and anticorrelated features had the same or the opposite disparity sign. All participants performed better when correlation and anticorrelation carried opposite disparity signs. (b) As a control, we measured performance when observers viewed purely correlated or anticorrelated stereograms. As expected, fully correlated stereograms were trivially discriminated by participants. However, observers performed at chance level for fully anticorrelated stereograms (two-tailed one sample t-test against chance performance, $t(11)=0.30$, $p=0.77$). Therefore, anticorrelated features alone did not elicit a reliable depth percept.

Specificity of the masking effect

In a series of follow-up experiments, we sought to characterize the specificity of the masking effect in space and time. First, we tested how similar the correlated and anticorrelated disparities should be for masking to be observed. We quantified masking as a function of the difference between the correlated and anticorrelated disparities (Fig. 3.3a), and found that masking varies as a function of the disparity separation between the correlated and anticorrelated dots (Fig. 3.3b). In particular, masking is maximal when correlated and anticorrelated dots have the same disparity, and there was little masking when they were separated by 6 arcmin of disparity. This suggests that the masking effect is narrowly tuned to the specific disparities conveyed by the anticorrelated portions of the display.

To probe the specificity of masking further, we relied on the fact that disparity channels in the human visual system are less tightly tuned for larger disparities[84,238–240]. In particular, we measured the disparity tuning of masking for different magnitudes of the disparity step-edge (Fig. 3.4a), obtaining one psychometric function for each

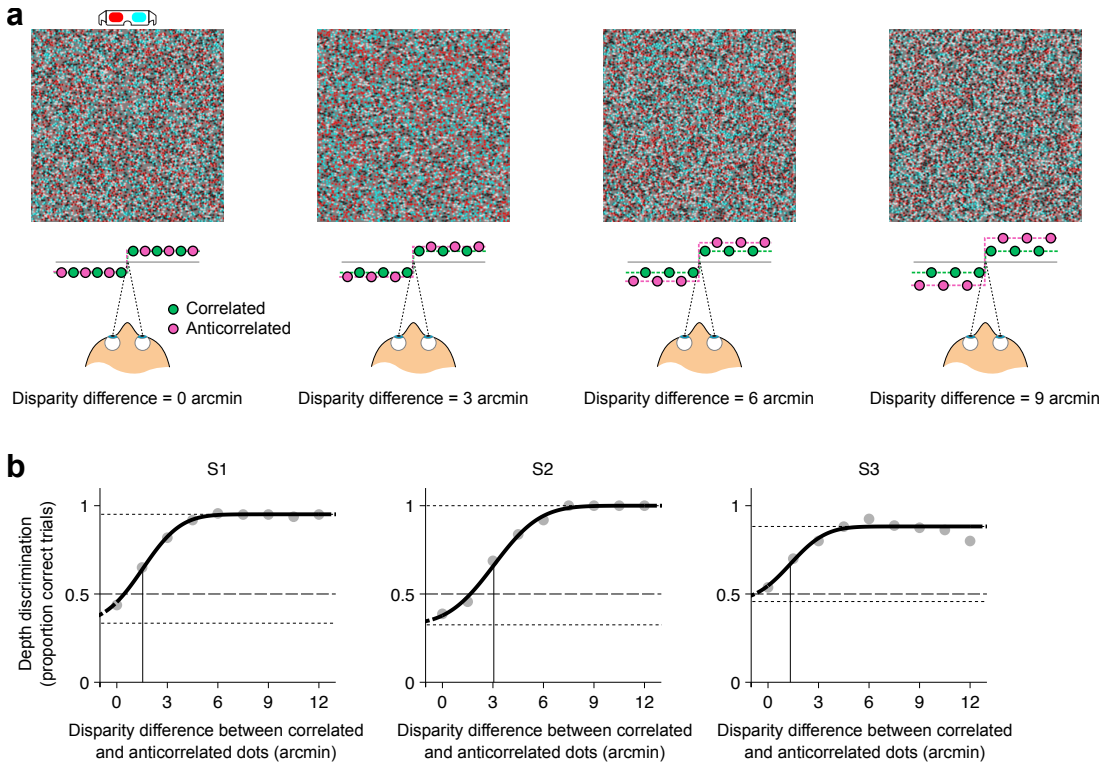


Figure 3.3: Specificity of masking by anticorrelation. The masking effect is reduced as the disparity difference between the correlated and anticorrelated elements increases. (a) Observers performed the step-edge depth discrimination task while we parametrically varied the difference between the disparities of correlated and anticorrelated dots. Stimulus illustrations designed to be viewed through red-cyan anaglyphs (red filter over left eye). (b) Proportion of correct choices as a function of the difference between correlated and anticorrelated disparities. Observers are at chance level when correlated and anticorrelated disparities carry the same disparity, but the masking effect gradually decreases — and eventually disappears for disparity differences greater that 6 minutes of arc. Circles denote proportion of correct choices for disparity differences ranging from 0 to 12 arcmin. Solid lines depict sigmoid curve fits (psignifit 3.0[236]).

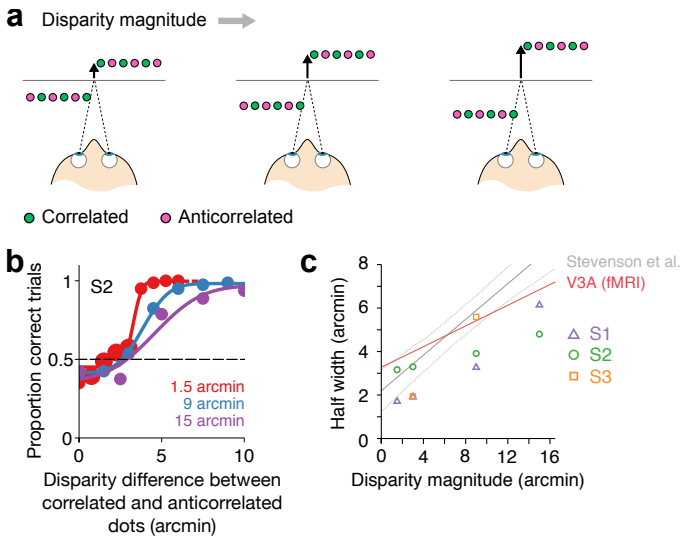


Figure 3.4: A size-disparity correlation for anticorrelation masking. (a) In addition to varying the distance between correlated and anticorrelated disparities, we also varied the disparity magnitude of the step-edge. (b) Example psychometric curves for different disparity magnitudes for one observer. As a proxy for specificity, we extracted a "half-width" parameter by finding the disparity difference at which the psychometric curve was half-way between its minimum and maximum. (c) We found that the effect tuning width, as indexed by the half-width parameter, increases with disparity magnitude. Our estimates (coloured symbols) line up well with previous psychophysical estimates of tuning width of disparity channels[240], as well as with fMRI-metric estimates from area V3A in the human brain[241]. Grey lines represent bootstrapped mean and 95% CI from reference[240]. Red line represents fMRI-metric estimates from reference[241].

magnitude (Fig. 3.4b). We quantified the specificity of masking using the disparity difference at which the observers' performance was half-way between its minimum and maximum value (analogous to the tuning half-width). We found that the tuning width of masking increased (quasi-) linearly with increasing disparity magnitude (Fig. 3.4c; Pearson's correlation, R=0.84, N=10, p=0.002; data pooled across subjects). Further, the tuning width values we obtained were remarkably consistent with previous psychophysical data[240] (Fig. 3.4c, grey lines) as well as estimates from fMRI recordings[241] (Fig. 3.4c, red lines).

Spatial extent of the masking effect

We have now seen that (i) the visual system uses disparity information contained in anticorrelated features, and (ii) it does so in a rather selective way. These findings describe the interaction between correlated and anticorrelated features in depth (i.e.

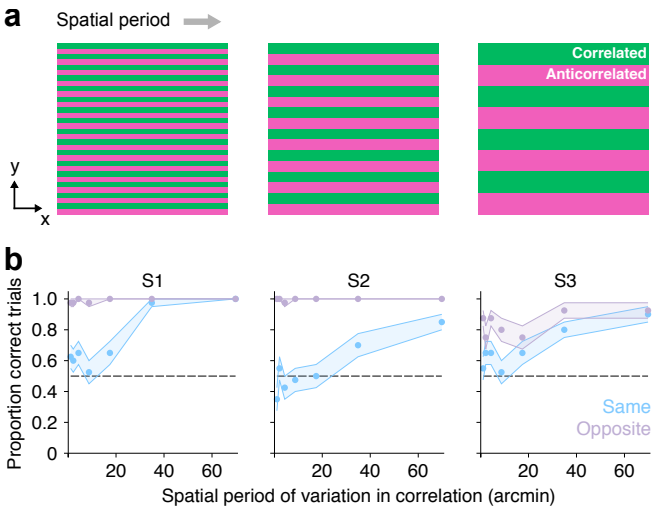


Figure 3.5: (a) Random-dot stereograms consisted of an even mixture of correlated and anticorrelated dots, but we varied their spatial arrangement along the vertical axis. The arrangement was dictated by a square-wave function with period ranging from 1 to 70 arcmin (seven equally spaced steps in logarithmic scale). This is depicted in cartoon form. (b) Proportion of correct responses as a function of the spatial period of the correlation square-wave for individual subjects. Blue elements depict performance for stimuli where correlated and anticorrelated disparities were equal. Purple elements depict data for stimuli where correlated and anticorrelated dots had disparities of opposite sign. The shaded area indicates bootstrapped 68% CI.

along the z-axis; Fig. 3.3a-3.4a). Next, we sought to characterize how the masking effect of anticorrelation depends on the visuotopic (i.e. frontoparallel) separation between correlated and anticorrelated features. We assessed this in a simple way by changing the spatial arrangement of correlated and anticorrelated dots in the display (Fig. 3.5a). In particular, we measured performance on the step-edge discrimination task when correlated and anticorrelated features were closely intermixed, or more widely separated. We found that the masking effect diminished when the correlated and anticorrelated features were spatially separated from each other (Fig. 3.5b), and had virtually disappeared for spatial periods greater than 30 arcmin. These data suggest that the anticorrelation masking effect might be limited by relatively small receptive field sizes, possibly in early visual cortex.

Asynchrony between correlation and anticorrelation

Having examined the spatial properties of masking, we now turn to its characterization in time. We recently suggested that neural responses to anticorrelated disparities

might be strongly mediated by suppression[231]. To implement this in cortex would entail the use of inhibitory interneurons, imposing an additional synapse (relative to excitation). This entails a temporal delay, and neurophysiological evidence suggests that delayed suppression might be an important mechanism in stereopsis[80,81]. Therefore, we postulated that manipulating the relative time of onset of correlated and anticorrelated dots in the display could modulate the masking effect. In particular, we predicted a stronger masking effect if anticorrelated features precede correlated features, as there is more opportunity to drive suppression before the onset of visual features that drive net excitation.

To test this idea, we manipulated the onset asynchrony between correlated and anticorrelated dots in the RDS, so that anticorrelated dots could lead or lag correlated dots by up to 33 milliseconds (Fig. 3.6a). While the relative timing of the correlated and anticorrelated portions of the stimuli was too short for observers to be aware of, we found that observers' performance was consistently worse when anticorrelated dots preceded correlated dots (Fig. 3.6b,c). This is compatible with the idea that anticorrelation drives suppression, and that suppression lags excitation.

A potential concern with this experiment is that anticorrelated features might disrupt binocular eye vergence, which, in turn, could alter performance on the depth discrimination task — particularly given our short stimulus presentation period (133 milliseconds). While this is plausible, our results do not support this view. First, the effect was observed even for very short asynchronies between correlation and anticorrelation. In the shortest asynchrony case, there was only 1 frame (i.e. 8.3 milliseconds) during which we displayed anticorrelated dots alone, which is unlikely to affect vergence. Second, the observers were highly trained to maintain vergence during stereoscopic viewing. Third, our stimuli were surrounded by a background pattern to promote a reference for stable vergence throughout the entire experiment. Fourth, we performed an additional experiment for one of the participants where we kept the onset asynchrony constant but we increased the total presentation time from 133 to 250 milliseconds, thereby allowing the subject to correct a putative transient disruption of vergence. The effect persisted, suggesting that the masking effect did not depend on the stimulus presentation time.

Finally, we tested whether the effect was specific to anticorrelation, rather than being a generalised metacontrast masking effect[242]. To do so, we included a control

condition in which anticorrelated dots in the stimulus were replaced by uncorrelated dots. Like anticorrelated features, uncorrelated dots do not support a coherent depth impression. However, under the proscriptive framework there is a critical difference in that uncorrelated dots do not drive suppression of a particular disparity. As such we did not expect differential masking from adding uncorrelated dots to the display. In line with this logic, we found that the observers' performance did not depend on the onset asynchrony between correlated and uncorrelated dots (Fig. 3.6d). This indicates that the temporal masking effect is specific to anticorrelation, which we interpret as a means of driving suppression of particular disparity signals.

3.4 Discussion

Our understanding of sensory processing has been guided by the loose intuition that computations and neural representations should resemble perceptual experience. Here we consider a case in which our perceptual experience of a world of solid objects (stereopsis) is enabled by neuronal responses that are selective for features that differ substantially from routine perceptual experience. In particular, we show that visual features that are unlike real objects in the environment, and which have long been understood as nuisance signals that complicate the interpretation of depth, in fact provide information that guides perception.

We tested human participants with stereograms engineered to provide either correlated or anticorrelated signals. Our logic was to test the effects of combining information for and against particular depth interpretations. Under the proscriptive framework, anticorrelation provides 'what not' information that drives suppression of depth values that are unlikely to have been evoked by the viewed scene. Here we demonstrate the perceptual consequences of doing this: depicting correlated and anticorrelated signals with the same disparity makes the correlated signal harder to see (Fig. 3.1e,f; Fig. 3.2). This masking effect is tuned to the disparity depicted by the anticorrelated signals, their spatial and temporal configuration. We suggest this reflects the effects of driving suppression of the disparity indicated by the anticorrelated elements. It is important to understand that these stimuli do not isolate correlated vs. anticorrelated signals. Rather, any viewed scene (including correlated RDS) will provide the brain with some visual features that are correlated and others that are anti-

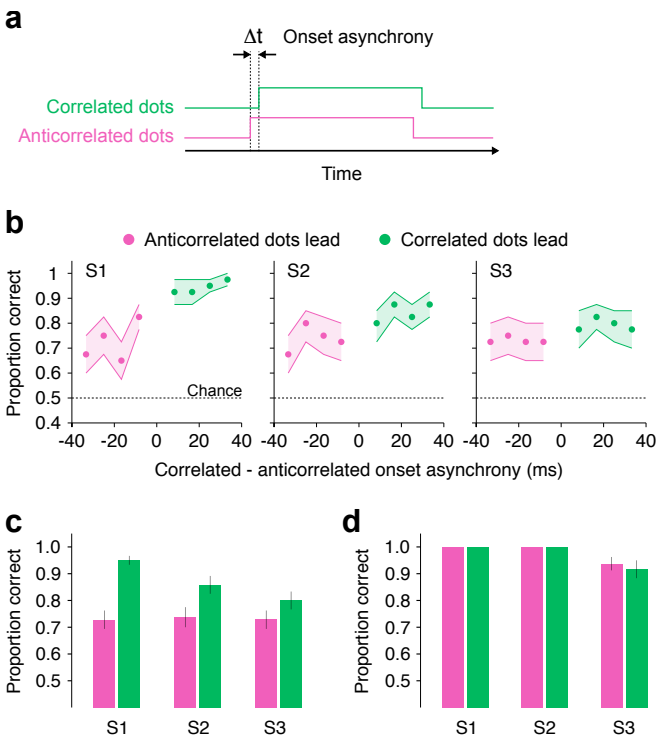


Figure 3.6: Effect of temporal asynchrony between correlation and anticorrelation. (a) We examined the perceptual consequences of varying the relative onset time of correlated and anticorrelated dots. Anticorrelated dots could lead or lag correlated dots by a temporal interval ranging from 8.3 milliseconds (1 monitor refresh frame) to up to 33.3 milliseconds (4 monitor refresh frames). (b) Proportion of correct choices was lower when anticorrelated dots preceded correlated dots (pink elements) compared to when correlated dots led (green elements). Points and shaded regions represent bootstrapped mean and 68% CI, respectively. (c) Proportion of correct responses for individual subjects obtained by grouping the data in panel b by leading dot type. Performance is worse when anticorrelated dots lead. (d) Proportion of correct responses for a control experiment in which we replaced anticorrelated dots by uncorrelated dots. Bar graphs and error bars represent bootstrapped mean and 68% CI, respectively.

correlated. The brain should use the totality of this information (i.e., both dissimilar and similar features) to estimate the most likely interpretation of the scene[231].

The perception of depth based on stimuli of inverted contrast polarity was first examined by Helmholtz[243]. Several studies have tested the perception evoked by anticorrelation, and suggested that these stimuli generally do support depth perception[150] except in very specific circumstances where a weak percept may emerge[16,33,34,37,38]. In the rare circumstances where depth perception emerges, the perceived depth sign varies. For instance, very sparse anticorrelated RDS can elicit a veridical disparity percept[33,34], whereas very large disparities were reported to elicit a weak and reversed depth percept[36]. It is known that the firing rates of disparity selective V1 neurons are reliably modulated by disparity in anticorrelated RDS[87,89]. This has suggested a discrepancy between perception and neurophysiology, with the interpretation being that the responses of V1 neurons to anticorrelated RDS are not closely linked to perception[33].

Our study demonstrates that the disparity carried by anticorrelated features is important for perception. We show that the visual system is exquisitely sensitive to differences in the disparity imposed in anticorrelated features (Fig. 3.3 and 3.4). Previous research concluded that correlation computations are negligible for fine disparities, since reversals of depth perception with anticorrelated RDS only occur for coarse disparities[36]. Our results challenge this idea. We show that the visual system is not blind to anticorrelated disparities, even when disparity magnitude of the step-edge was small (1.5 arcmin) (Fig 3.4b, c).

Additionally, we show that the effects of anticorrelation are more specific for small disparity magnitudes than they are for large disparity magnitudes (Fig. 3.4c). These properties are usually attributed to the detection of similar features in the two eyes[84,239,240]. Such similarities suggest the existence of common underlying neural machinery responsible for sensing a continuum of binocular correlation: from highly similar to highly dissimilar. Could the masking effect of anticorrelation be explained by residual perception of depth based on anticorrelated features alone? Our control experiments do not suggest this is the case — observers performed at chance level when we replayed the step-edge stimuli rendered exclusively with anticorrelated dots (Fig. 3.2c). Additionally, veridical or reversed depth perception based on anticorrelated features alone would predict differences in performance when the disparity in

anticorrelated and correlated dots deviate by the same magnitude, but in opposite directions. By contrast, we found that the perceptual tuning curves (e.g. Fig. 3.4b) for opposing directions were strongly correlated ($R=0.81$, $N=100$, $p<0.001$). Together, this suggests that residual perception based on anticorrelated features alone does not explain our results.

It is tempting to speculate about the extent to which models of disparity selectivity in V1 can account for these results. However, the link is not straightforward. The energy model can account for both veridical and reversed depth "percepts" for anticorrelated stereograms, depending on the particular parametrization of the energy model[186,244]. This is also the case for the binocular likelihood model[231]. As others have concluded[33,107], additional processing of stereoscopic information that supports perception likely takes place in extrastriate visual cortex.

It is important to note that we found considerable between-subject differences in the extent to which individuals could tolerate anticorrelated elements added to the display (Fig. 3.2a; i.e., note the absolute position of each participant's data on the ordinate axis, rather than the relative difference between the two conditions). While we found that the contrast between Figures 3.1e and 3.1f works well for 90% of the two hundred or more people we have tried this demonstration on, quantifying this effect in a laboratory setting clearly challenges participants' visual systems. Experienced psychophysical observers could tolerate enough anticorrelation added to the stimuli to give us sufficient experimental dynamic range to explore the effects in detail. We know that training can have profound effects on stereoscopic vision[245,246]. It is possible that the suppression produced from anticorrelation is so strong in naive participants that it becomes difficult to tolerate even a small amount, especially under the brief presentation durations needed for well-controlled psychophysical testing.

The neural computations that support stereoscopic vision likely involve suppression[82,89,209,231,247], especially for anticorrelated disparities[231]. Neurophysiological data suggest that suppressive input onto disparity detectors is delayed with respect to excitation[80]. Thus, we hypothesize that the response to anticorrelated stimuli may be partly mediated by delayed suppression. In agreement with this, we show that 'front-loading' the visual system with anticorrelated dots, putatively eliminating the delay between excitation and suppression, has a greater effect on performance than when anticorrelated dots are presented after correlated dots (Fig. 3.6b,c). More generally,

our data suggest that suppression shapes the tuning of disparity detectors — an idea that has received strong neurophysiological support[80,81,214].

Humans can detect variations in disparity over spatial regions as small as 4-8 arcmin[248,249], a resolution measure that has been linked to the size of V1 receptive fields[250]. Although consistent with the size of V1 receptive fields, our results suggest that disparity in anticorrelated features might be pooled over a somewhat greater extent. We observed strong masking from anticorrelation for signals separated by up to 20 arcmin (Fig. 3.5b). As a result, it is possible that the interaction between disparity in correlated and anticorrelated features spans a greater extent than the interaction between disparity in correlated features. This would point towards a form of extra-classical contextual modulation effect, perhaps implemented by means of a centre-surround-like receptive field organization for correlated and anticorrelated disparities. However, further psychophysical testing is needed to directly answer this question. To our knowledge, neurophysiological investigations have not yet addressed this, but recent work suggests that complex cells rely on extensive pooling[77,79].

Neural responses to anticorrelated visual features have long appeared surprising: a physical feature in the environment will not project anticorrelated intensities to the two eyes, and viewing anticorrelated displays does not support a robust perceptual interpretation. As such, these signals were thought to complicate the perceptual interpretation of depth. By contrast, here we show that they guide perception: using "what not" information to suppress unlikely interpretations of the viewed scene[231]. Our experimental paradigm relies on showing how features that we "cannot see" can "hurt" perceptual judgments. However, the general mechanism that we reveal is central to routine perceptual judgments: by combining evidence for and against the likely structure of the scene, the brain can make best use of neural responses tuned to the types of inputs experienced in everyday life[231].

Chapter 4

Cortical Organization of Binocular Disparity in Human Visual Area V3A

This chapter reproduces the work associated with the following published manuscript:

Goncalves NG *et al.*. 7 Tesla fMRI Reveals Systematic Functional Organization for Binocular Disparity in Dorsal Visual Cortex. *Journal of Neuroscience*, 35(7), 3056–3072. 2015.

4.1 Introduction

Understanding cortical organization at the mesoscopic scale is an important step in characterizing local neural circuits. The hypercolumn concept[138] has been extremely influential in modeling cortical functional architecture[251], and provides a framework to test the neural machinery that facilitates sensory processing. However, despite good knowledge from animal models, comparatively little is known about the local architecture of human cortex.

Here we investigate the brain organization that may underlie our ability to make precise depth judgments based on binocular disparities. Extracting depth from disparity is a demanding, yet routine, neural computation, making it plausible that it is supported by systematically organized cortical structures. Recordings from cat and macaque cortex indicate that neighborhoods respond similarly to disparity: i.e., electrophysiology and optical imaging showed that disparity populations are structured

in V2[145–147,252,253], V3/V3A[98,148,149] and MT/V5[100]. By contrast, area V1 is reported to show, at best, weak clustering[143,144].

Tests of disparity processing in the primate brain point to strong fMRI responses in area V3A[51,129,193]. This work complements evidence that macaque V3A contains clusters of disparity selective neurons[98,148,149]. Here we therefore focus on responses to disparity in the dorsomedial region around V3A using human brain imaging. To benefit from improved signal-to-noise and high BOLD contrast-to-noise ratios[254], we used ultra-high field (UHF) 7 T fMRI. Recent UHF fMRI work indicates that it can link structures observed in animal models with representations in the human brain: organization for ocular dominance and orientation was reported in primary visual cortex[255,256], while structured responses to motion direction were observed in area MT/V5[257].

We report that human visual cortex is systematically organized for binocular disparity, with dorsomedial area (V3A/B) showing structure that relates to the functional characteristics of depth perception. First, we test for clustering of disparity response profiles, finding evidence for maps that are reproducible across imaging sessions. We then characterize the selectivity of individual voxels for disparities of different magnitudes. We find different profiles of voxel responses: some show fine-tuned responses while others have categorical responses (i.e. near vs. far depth). By fitting Gabor models to voxel profiles, we show a relationship between the magnitude of disparity and the tuning width of voxels in V3A and V3B/KO, but not earlier visual areas. Finally, we demonstrate a similarity between these voxel responses and established models of the functional properties of human stereopsis. Together, our findings suggest that dorsal visual cortex (V3A, V3B/KO) contains specialized organization for disparity, which may support the neural computations underlying 3D perception.

4.2 Materials and methods

Participants

Six subjects (three male, aged 25 – 38 years, including authors N.G., H.B., A.E.W.) participated in the study. Participants provided informed consent and procedures

were approved by the University of Nottingham Medical School Ethics Committee. All participants had normal or corrected-to-normal vision and did not present stereo deficits. One (Participant 5) withdrew from the study following the second scan.

Stimuli and design

Stimuli were presented stereoscopically using red and green anaglyphs (the very tight confines of the head coil meant that other stereoscopic display techniques were not feasible). Participants viewed stimuli projected onto a screen located at their feet (viewing distance = 242 cm). To view the screen from within the head coil, they wore prism glasses (to which the red and green filters of the anaglyphs were attached). Stimuli were rear-projected onto the display screen from an EPSON EMP-8300NL using a Nivitar NuView long-throw lens. At the start of each scan session, we verified the correspondence between disparity sign (e.g. 'negative') in software with that presented on the projection screen (i.e., perceptually 'near'): an experimenter viewed the screen from the front while wearing the prism glasses with anaglyph filters attached. Stimuli consisted of random dot stereograms (7 x 7 deg) on a mid-gray background, surrounded by a static grid of black and white squares intended to facilitate stable vergence. Dots in the stereogram followed a black or white Gaussian luminance profile, subtending 0.07 deg at half maximum. There were 108 dots/deg2, resulting in around 38% coverage of the background. In the center of the stereogram, four wedges were equally distributed around a circular aperture (1.2 deg), each subtending 3 degrees in the radial direction and 70 degrees in polar angle, with a 20 degrees gap between wedges (Fig. 4.1A). We varied the depth of the wedges by modulating disparity levels in relation to the fixation point (3, 9, 12, 15, 24 and 36 arcmin, ± 0.5 arcmin jitter, crossed and uncrossed). At a given time point, all wedges presented the same disparity. To reduce adaptation, we applied a random polar rotation to the set of wedges such that the disparity edges of the stimuli were in different locations for each stimulus presentation (i.e., a rigid body rotation of the four depth wedges together around the fixation point). In the center of the wedge field, we presented a fixation square (side length = 1 deg) paired with horizontal and vertical nonius lines. Four participants underwent three imaging sessions (Participants 1, 2, 3 and 6), while the remaining participants (Participants 4 and 5) took part in two sessions (sessions 1 and 2). These

imaging sessions were performed on different days. In sessions one and two, we presented stimuli at fine-to-intermediate disparity levels (±3, 9 and 15 arcmin). In the third session, we delivered stimuli at intermediate-to-coarse disparities (±12, 24 and 36 arcmin). BOLD responses to binocular disparity were estimated using a block design. During each block, stimuli were presented at one of the six disparity levels defined for that session. The block length was 15 seconds, with 10 stimuli presented for 1 second with an inter-stimulus interval of 0.5 seconds. During a run, six different blocks were presented (one for each disparity level), and each was repeated three times (18 blocks). In addition, there was a fixation block at the start and the end of each run. Each run lasted 300 seconds (20 blocks x 15 s), and we collected eight or nine functional runs in each imaging session. On each run, we asked participants to fixate in the central fixation square while performing a Vernier detection task[193].

Imaging

Imaging sessions were performed at the Sir Peter Mansfield Magnetic Resonance Centre, University of Nottingham on a 7 Tesla Philips Achieva scanner with volume transmit and a 32-channel receive coil. Head motion was restricted by the use of foam padding and a vacuum pillow (B.u.W. Schmidt). Data were acquired using a three-dimensional gradient echo echo-planar imaging (3D GE-EPI with SENSE factor 2.35 in the anterior-posterior (AP) direction and 2 in the foot-head (FH) direction, TE/TR = 28 / 82 ms, FA = 22 deg, EPI factor 45; 0.96 x 0.96 x 1 mm3; Matrix size: 160 x 160 x 36 (AP x RL x FH); volume acquisition time of 3 s; 100 volumes per run) to acquire blood oxygen level-dependent signals from a field of view (FOV) spanning the dorsomedial visual cortex. The reduced FOV required the use of outer-volume suppression in the phase encoding direction (AP) to prevent signal fold-over. Prior to 7 T imaging sessions, participants underwent localizer scans at the Birmingham University Imaging Centre (BUIC). A 3 Tesla Philips Achieva scanner was used to collect fMRI data to standard retinotopy experiments[193]. Retinotopic maps were later used to define regions of interest, as well as to assist in positioning the acquisition volume over dorsomedial visual areas for 7 T data acquisition. An anatomical volume was also acquired (MPRAGE, 1mm isotropic resolution) and used for surface reconstruction using FreeSurfer[258,259]. The resulting reconstructed

white-matter (WM) and gray-matter (GM) surfaces were then used to compute cortical profiles. These cortical profiles were defined as vectors that connect corresponding vertices in WM and GM surfaces. These vectors were then used to sample functional data at different relative depths. We analyzed functional data using mrTools (`http://www.cns.nyu.edu/heegerlab`) and custom Matlab code (The Mathworks Inc, Natick, MA). We first co-registered functional scans to the anatomical volume used for surface reconstruction, and subsequent analyses were performed in the individual native space. Preprocessing consisted of motion correction using linear interpolation and linear detrending. We modeled the BOLD signal using the six disparity levels as regressors of interest, and estimated the model parameters (i.e. the beta-weights) associated with each condition. Voxel preference was assigned on a 'winner-take-all' basis with respect to the magnitude of the beta-weights across conditions. Voxel disparity preferences were mapped onto the cortical surface by sampling the functional data (nearest neighbor interpolation) at intermediate depths along a cortical profile. Although measurements at ultra-high field are less susceptible to vascular influences[260–262], sampling at intermediate depths further minimizes the influence of superficial veins[263] and improves spatial localization[264]. Nevertheless, we also quantified vascular influences by calculating the mean BOLD signal amplitude across the cortical surface.

Multivoxel pattern analysis

In order to confirm that our target regions carried information about the stimulus dimension that we manipulated, we employed standard multivariate analyses on activity from retinotopically-defined areas in dorsomedial visual cortex[193]. For each region-of-interest, we converted voxel time series to z-scores, and shifted the respective time course by two volumes (equivalent to six seconds) to account for the hemodynamic delay. We then averaged data-points within each condition block, and used a linear classifier (support vector machine, libsvm toolbox[265]) to discriminate between different stimulus conditions. We ranked voxels according to the z-score of the comparison between all stimuli and the fixation blocks, and then used the top 500 voxels in each ROI for the classification analysis. We followed a leave-one-out cross-validation procedure, resulting in eight or nine folds, depending on the number of completed

runs for each participant (we use the term 'fold' to refer to the different combinations of independent subsets of the data). In particular, from seven/eight runs (out of the total eight/nine runs) we extracted 126/144 patterns to train the classifier and then tested the classifier on 18 patterns extracted from the remaining run. This process was repeated so as to leave out each individual run in turn, and the mean accuracy for each subject was computed across folds. We performed two-way (near versus far) and 6-way (individual disparities) decoding analysis using this technique.

Calculating the probability of similar voxel preferences in a local neighborhood

To assess the degree of clustering in responses to disparity, we examined the distribution of disparity preferences in the local neighborhood of voxels surrounding a given target voxel. We first sub-divided the data into two independent sets (one to find the disparity preference of the target, the other to find the preference of the surround), using a leave-two-runs out cross-validation procedure. For each individual voxel, the disparity preference was estimated by fitting a General Linear Model (GLM) to six/seven runs out of the total of eight or nine runs acquired. We used the remaining two runs to estimate the disparity preferences of the voxels adjacent to the target voxel. Voxels were considered neighbors if (i) they belonged to the 26-connected neighborhood that shared a vertex with the target voxel, and if (ii) they were located within the region of interest (e.g., V1) under consideration. We calculated the frequency of each disparity preference in the neighborhood of individual voxels, and indexed the distribution to the disparity preference of the target voxel. We repeated this process using different subdivisions of the data for cross-validation (${}_8C_2 = 28$ or ${}_9C_2 = 36$ folds, depending on the number of experimental runs acquired) for each participant, and pooled the resulting frequency distribution across subjects. Frequencies were converted to probabilities by dividing by the total number of adjacent neighbors, and then averaged according to the preference of the central (target) voxel. This produced six probability distributions that describe disparity preferences in the surround of individual voxels (one distribution per central disparity preference). To compensate for general biases in disparity preference, we divided each probability distribution by the overall disparity preference probability within that region of interest. These six

relative probability distributions are represented in matrix form, where each distribution is represented by a row-vector. Rows indicate the (indexed) central preference, and columns represent the local disparity preference.

Simulating columnar organization and local clustering

To quantify the extent to which disparity clustering at the neural level might reasonably be extracted by a coarser-scale sampling grid (i.e., fMRI voxels), we simulated cortical architectures with different spatial scales and then used the clustering analysis described in the previous section. In particular, we simulated cortical columns of different spatial periodicity by bandpass filtering two-dimensional white noise[266], a method that has previously been used to simulate orientation columns[267]. We started by generating a matrix whose elements where pseudo-randomly extracted from a normal distribution, representing a 40 mm^2 patch of cortex. The noise matrix was bandpass filtered to preserve content with a specific periodicity, which determines the columnar width of the pattern. To test different levels of clustering, we simulated cortical columns varying in width between 1 and 4 mm. Having generated the neural map, we simulated the fMRI sampling procedure by placing a two-dimensional grid of 1mm squares (representing voxels) over the columnar map. We then assigned a preference to each 'voxel' using a probabilistic approach. In particular, we defined the probability of voxel preference across trials as the distribution of the underlying neural preferences within each voxel. This provided us with a discrete probability distribution for each voxel, which we then used to generate 500 fMRI preference maps. We then assessed local disparity clustering as above (see Calculating the probability of similar voxel preferences in a local neighborhood). The only difference was that we considered the 8-connected neighborhood of each voxel, as the simulation was performed using a two-dimensional representation, rather than the three-dimensional data obtained from our empirical measurements.

Comparison of disparity preference maps across sessions

To determine whether disparity preferences revealed at the voxel level represented a stable property of cortical responses, we sought to compare disparity maps obtained from scans performed on different days. To this end, we first needed to identify those

voxels that had reliable disparity responses within each session. We did this by estimating the disparity response of each voxel using a leave-two-runs-out GLM fitting approach. By iteratively leaving two runs out, we identified voxels that responded maximally to a given disparity on at least 50% of the GLM fits. Having identified voxels with stable within-session responses, we re-estimated the disparity response of each voxel using the full dataset (i.e., a GLM fit to all runs within a session). To co-register maps from different sessions into a common space, we transformed measurements from each participants' original functional space to their native anatomical space by applying the transformation matrix computed during anatomical-functional co-registration. We then computed the Pearson correlation between corresponding voxels (nearest neighbors) across sessions (bootstrapping, 10,000 samples). In order to ensure the stability of the correlations, we systematically varied the within-session repeatability criterion (from 50 to 80% of the same preference using the leave-two-runs-out GLM procedure). We found that our estimates of between-session correlations were stable across this range of within-session repeatability thresholds. As the co-registration procedure described above is not perfect, we also used an additional alignment step to compensate for small misalignments between data acquired in different sessions. In particular, we recomputed correlations between voxels in different sessions after applying an additional iterative alignment procedure to improve co-registration. This procedure adjusted the position of one of the maps, so as to minimize the differences in disparity preference across the region of interest (we provide results with and without this extra alignment procedure). We defined the first session map as the reference and the second session map as the source. Let R and S represent the disparity preferences of the reference and source maps, respectively. Each of these is defined as an m-by-four matrix, where m is the number of voxels, and each voxel is described by four features: their three-dimensional coordinates (x, y, z) and their peak disparity response. We iteratively adjusted an affine transformation W to the source map S so as to minimize the disparity preference difference between corresponding nearest neighbors across sessions, d, which we can define as

$$d = \sum_i ||R_i - (WS)_{c(i)}|| \tag{4.1}$$

where $c(i)$ is the index of the closest voxel of WS in relation to R_i. We restricted

the transformation W to be as small as possible by penalizing large deformations. We thus defined our optimization function as

$$J = d + \lambda||W - I|| \tag{4.2}$$

where I is the identity matrix and lambda is an empirically-defined regularization weight equal to 0.1 that ensured convergence of the minimization algorithm while preventing gross distortions of the maps. The first term of the optimization objective minimizes overall differences in spatial organization of disparity preferences between maps, while the second term restricts the spatial transformation to be as small as possible, weighted by lambda. The maximum number of iterations was set to 200 and the resulting spatial transformations were very close to identity. After this alignment step, we recomputed the Pearson correlation coefficient using bootstrapping (10,000 samples, as above).

So far we have used the Pearson correlation coefficient to quantify the correspondence between disparity maps obtained in different imaging sessions. This is based on a linear relationship between variables. A more general approach is to ask how much information is shared between disparity maps obtained in different imaging sessions. To do so, we used mutual information[196], which quantifies the reduction in uncertainty about a variable after the observation of another variable. In particular, we computed the reduction in uncertainty about disparity preference in one map, after observing the disparity preferences in the other map (note, we performed this analysis on the data without the additional preference alignment step). In the discrete case, mutual information is defined as[268]

$$I(X;Y) = \sum_{(x,y)} P_{XY}(x,y) log \left(\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} \right) \tag{4.3}$$

where $P_{XY}(x,y)$ denotes the joint probability distribution of X and Y, while $P_X(x)$ and $P_Y(y)$ represent the respective marginal probability distributions. If X and Y are independent, $P_{XY}(x,y) = P_X(x)P_Y(y)$, and therefore $I(X;Y) = 0$.

Modeling disparity responses of individual voxels using neuronal templates

To model the responses of individual voxels to different disparities, we used the disparity tuning templates proposed by Poggio[58,83]. In particular, we used linear regression to assess how tuned (TN – tuned near / TF – tuned far), categorical (NE – near / FA – far) and excitatory/inhibitory (TE – tuned excitatory / TI – tuned inhibitory) cell models explained individual voxel responses. Regressors consisted of discrete ideal responses for each model type at the preferred disparity of that voxel and an additional offset/baseline term. Specifically, the discrete realizations of these models were: (i) tuned model — a Kronecker delta shifted to the preferred disparity of the voxel, (ii) categorical model — a square wave cycle between 0 and 1, odd around zero disparity, so that the positive step coincides with the preferred disparity of the voxel, and (iii) excitatory/inhibitory model — a shifted triangle wave cycle, even around zero disparity. The triangle wave had its peak around zero disparity if the disparity preference of the voxel was the smallest disparity magnitude presented, and its trough otherwise. After assembling the regressors for each voxel, linear regression was performed using Matlab. This produced a set of four weights per voxel (one for each tuning model, plus the baseline term), which express the extent to which each model explained the response profile of individual voxels. For subsequent analysis, we selected voxels that were well modeled by this approach ($R^2 > 0.8$).

Modeling voxel responses to disparity using a Gabor model

The previous section used descriptive neuronal models to examine voxel responses. Next, we sought to estimate disparity responses more parametrically. To this end, we used a one-dimensional Gabor model that has been used to describe the response profiles of disparity selective neurons in early and extrastriate visual areas (e.g. V1[73]; V3/V3A[98]; MT[269]). In particular, we used a Gabor function to describe the response of voxels to variations of binocular disparity. For each voxel, we started by removing baseline differences in beta-weights by subtracting the mean beta-weight across all of the presented disparities. For each region of interest, we then grouped voxels based on their preferred disparity (i.e., maximum beta-weight), resulting in sixty groups (twelve preferred disparities for five regions of interest). We then fit a Gabor model

to each group of voxels (using the data from all the voxels, rather than the averaged voxel response), where the response to a disparity, d, was defined as

$$G(d) = A_0 + A\exp^{(-(d-d_0)^2/(2\sigma^2))} cos(2\pi f(d-d_0)+\phi) \tag{4.4}$$

where A is the amplitude, A_0 is the baseline, d_0 is the position of the Gaussian envelope, σ is the width of the envelope, f is the frequency of the cosine and ϕ is the phase shift between the cosine and the center of the Gaussian envelope. We used constrained optimization (fmincon, MATLAB) to find the parameters of the Gabor model that best described each group of voxel responses (least-squares estimation). We constrained the minimizers ad hoc to sensible values given the disparity levels we presented, which ranged from —36 to 36 arcmin. First, as baseline correction was performed prior to fitting, we constrained the baseline shift to values between —1 and 1. Second, as voxels were grouped by their preferred disparity prior to fitting, the position of the Gaussian was constrained to a window of 10 arcmin around the preferred disparity of each voxel group. Third, the amplitude of the Gaussian envelope was constrained to 1.2 times the amplitude range of voxel responses, while the width of the Gaussian was restricted between 5 to 12 arcmin to avoid overfitting the data. Finally, the frequency was allowed to vary between 0 and $1/(dmax-dmin)$ cycles per arcmin, where $dmax$ and $dmin$ represent the maximum and minimum disparity presented during the experimental session. This constrained the frequency to remain below half the sampling frequency. Using these parameter limits enabled us to avoid gross over-fitting that can arise from the oscillatory term of the Gabor (a combination of envelope width and frequency of the carrier). We quantified over-fitting by contrasting the Gabor model fit against a piece-wise linear fit to neighboring points in the response profile. In particular, we started by estimating the slope of the line that connects two consecutive points of the response profile. Then, we computed the maximum instantaneous variation of the fitted Gabor within the same interval, and subtracted the slope of the linear fit. If the Gabor oscillates considerably between two consecutive points, there will be a considerable absolute difference between the maximum variation of the Gabor and the slope computed by linear approximation. By contrast, if the Gabor follows the linear trajectory between two consecutive points closely, this difference will be nearly zero. Using the constraints described above, the

optimization function found good fits across experimental conditions. Specifically, we assessed the quality of fits using a χ^2 goodness-of-fit test[269]. This test compares the variance of the residuals around the mean tuning profile with the variance of the residuals around the model fit. In particular, we computed the difference between each data point and the model value at that disparity, which provided us with a distribution of residuals around the model. We then compared the variance of this distribution against the variance of the residuals around the mean using a χ^2 test for equal variances. The fit is considered satisfactory if the variances of these distributions do not differ significantly. The constraints of fMRI data acquisition meant that we were limited in the number of different disparities that we could measure reliably during each imaging session. In consequence, fits to voxel responses are limited in their resolution along the disparity domain. This presents a challenge in choosing and fitting the correct model to the data. Based on the electrophysiological literature, a Gabor model is a good descriptor of individual neuron responses within the visual cortex (see above). As an alternative we also considered a Gaussian model that has the advantage of fewer parameters. However, comparison of the models indicated that the Gaussian model was insufficient to capture the different profiles of the voxels we measured. In particular, thirty (out of a total of sixty) fits did not pass the χ^2 goodness-of-fit test described above ($p < 0.05$). By contrast, using the Gabor model only four out of sixty fits failed this test. This is not surprising since the Gabor model has more free parameters than the Gaussian. Therefore, we also compared these models using the Akaike information criterion (AIC), and found that the mean AIC value across all fits was much lower for the Gabor model. We therefore adopted the Gabor model to describe the response profiles of our sampled voxels.

Using fMRI-based estimates to model disparity population characteristics

The modeling methods so far described allow us to describe properties of disparity selective populations based on our fMRI recordings. Next, we investigated whether these estimates could be related to the characteristics of neural populations that underlie psychophysical depth judgments. Specifically, modeling and psychophysical investigations point to a relationship between disparity selectivity and disparity mag-

nitude — as disparity magnitude increases, disparity detectors are thought to have larger receptive field sizes[84,240], and this relationship is well approximated by a linear increase for disparities (5–20 arcmin) near fixation[240]. Motivated by these findings, we investigated whether the envelope size of the fitted Gabor models increased with disparity magnitude. In particular, we examined whether there is a correlation between the standard deviation σ and preferred disparity in each region of interest (Pearson's correlation, $p < 0.05$). If the correlation was significant, we used linear regression to estimate the best fitting trend that describes the variation of each Gabor parameter as a function of disparity magnitude; otherwise, the Gabor parameters were assumed to be constant and set to the mean value across disparity magnitudes. We centered this analysis on the relationship between the standard deviation of the Gaussian envelope and disparity magnitude since we were interested in changes in response profile width. The parameters of the cosine term — i.e. the frequency f and phase ϕ — provide insight into disparity selectivity in terms of the presence of on-off sub-regions within a neuron's receptive field. However, in our case the limited number of disparities sampled, and the aggregated nature of the voxel measurements meant that it would be difficult to draw any strong conclusions from the any observed relationship between these parameters; we therefore limited our analysis to the relationship between the peak response and standard deviation.

Using the estimates of the relationship between disparity magnitude and envelope size, we built a distributed population of disparity selective units. (Here, the term 'unit' describes a disparity detector, which, in this case, is derived from a population of voxels). For each region of interest, the regression fits were used to build Gabor detector units at seventeen equally spaced disparity magnitudes between −40 to 40 arcmin. We then simulated the ability of this bank of detectors to discriminate different disparities[84]. In short, we estimated the smallest disparity difference that could elicit a significant change in activity across the population as a whole. First, we computed the responses of each Gabor unit to two disparity levels, and derived the respective variances assuming direct proportionality. Specifically, if we let R_{ij} be the response of unit i to stimulus j, its variance is then given by $\sigma_{ij}^2 = kR_{ij}$ with $k = 1.5$ (for plausibility of this arbitrary parameter, see[84]). The number of standard

deviations separating these responses was defined as

$$d_i' = \frac{|R_{i1} - R_{i2}|}{\sqrt{(\sigma_{i1}^2 + \sigma_{i2}^2)}} \tag{4.5}$$

Large values of d_i' suggest that changes in response of the i^{th} unit were stimulus-induced, whereas small values indicate chance fluctuations due to noise. Statistically, the probability of observing a stimulus-induced change in individual units was defined as

$$p_i = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{d'} e^{-x/2} dx - 1 \tag{4.6}$$

At the population level, however, each unit represents one of many dimensions. In this multidimensional space, assuming uncorrelated noise, variation in responses were tested using the joint probability

$$p = 1 - \prod_{i=1}^{N} 1 - p_i \tag{4.7}$$

Here, p represents the probability of changes across the whole population being stimulus-induced. We finally computed the disparity discrimination threshold as the minimum disparity difference for which $p > 0.5$ (as in[84]; see their Erratum). Discrimination thresholds were evaluated at disparities between −40 to +40 arcmin.

4.3 Results

We presented participants with disparity-defined wedges at a range of different depth positions (Fig. 4.1A,B) and recorded the blood oxygenation level-dependent (BOLD) signal from voxels spanning the dorsomedial visual cortex (Fig. 4.1C). We observed strong BOLD responses to manipulations of disparity that were well localized to the gray matter (Fig. 4.1D). To provide a first analysis of these data, we quantified aggregated BOLD responses in the dorsal visual cortex using two approaches. First, we computed the change in the BOLD response for disparity-defined stimuli relative to the fixation baseline in each of the localized regions of interest. This revealed large

changes in the BOLD signal in V1 and dorsal extra-striate cortex (Fig. 4.1E), with magnitudes consistent with previous studies at ultra-high field[254,264,270]. Second, we quantified responses in different regions of interest using a multivoxel decoding analysis approach for disparity-defined stimuli[193]. In particular, we calculated the accuracy of a support vector machine in predicting whether a stimulus was nearer or farther than the fixation point based on patterns of voxel activity. We found high accuracies for discriminating crossed ('near') vs. uncrossed ('far') disparity (Fig. 4.1F) that were, on average, highest in V3A and at accuracy levels comparable with previous work[193]. We also performed a six-way classification analysis of the data, testing how well the presented disparity could be predicted from the six different types of stimuli (i.e., disparity values) presented. We observed performance well above chance, with highest mean performance in V3A (Fig. 4.1G). These results are consistent with previous work at 3 T in suggesting strong responses to disparity, particularly in areas V3A and V3B/KO[51,129,193]. This initial examination of the data is confirmatory. However, our primary interest was not in aggregated voxel responses from within different regions of interest, but rather whether 7 T fMRI would allow us to detect and quantify consistent spatial organization of individual voxel responses. In our next analyses, we therefore move to consider the response profiles of individual voxels as a proxy that summarizes the activity of a neural population centered on the voxel. To this end, we used the beta-weights of the GLM model fit to the fMRI time series, to determine how the different presented stimuli explain the activity of individual voxels. We start by defining the disparity preference of a voxel as the condition that yields the highest beta-weight (i.e. winner-take-all labeling). Later, we consider other models that seek to capture the response profile of a voxel based on all the estimated beta-weights.

The spatial clustering of disparity preferences

Motivated by reports of disparity clustering in macaque extrastriate cortex[98,149], we tested for clustering within the human visual cortex. In particular, we examined the spatial distribution of disparity preferences across the cortical surface by labeling individual voxels according to the disparity value that evoked the highest level of fMRI activity (i.e., maximum beta-weight of the GLM) during each imaging session. To visualize the data, we color-coded the disparity preferences of individual vox-

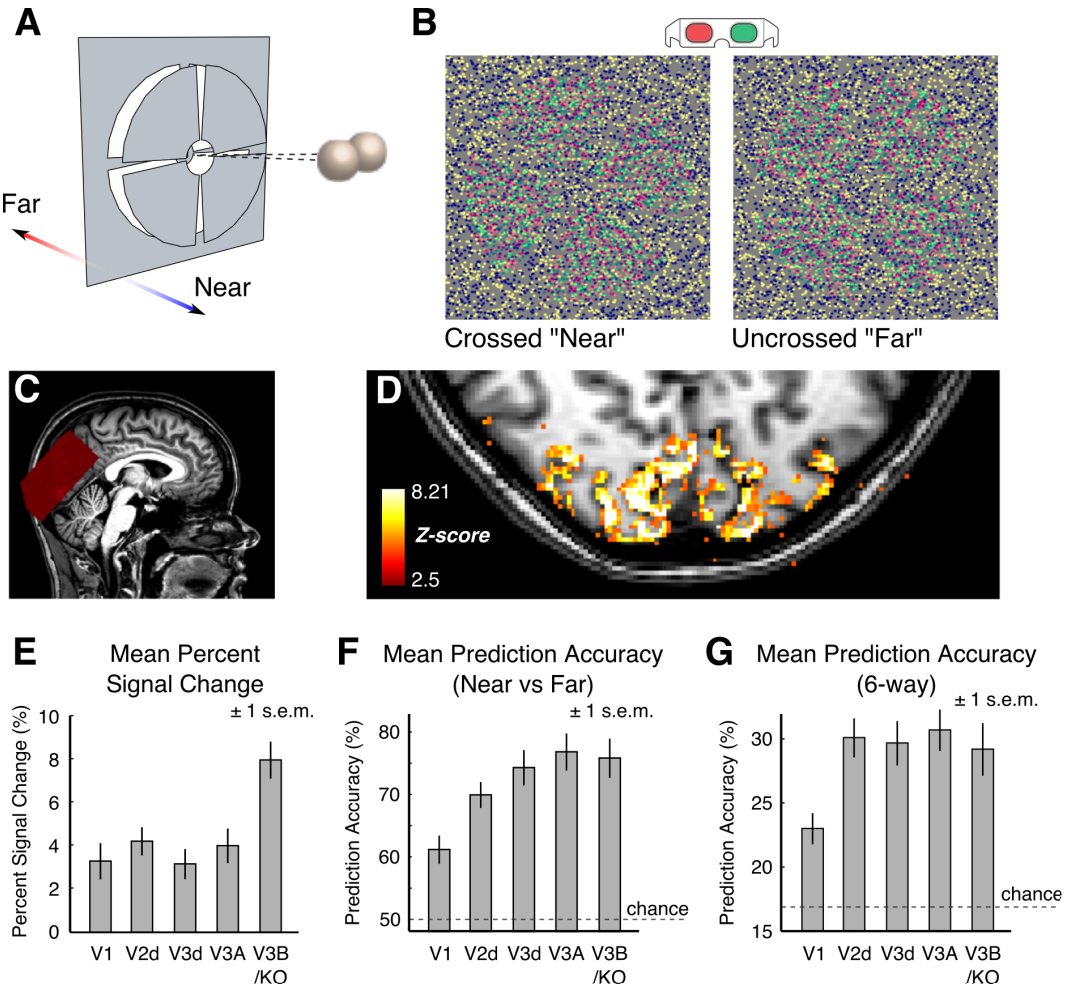


Figure 4.1: Schematic illustration of the stimuli and basic functional activations. A, Diagram of the depth arrangement in the stimuli. Four disparity-defined wedges were simultaneously presented at one of six disparity-defined depths during each imaging session (± 3, 9 and 15 arcmin in sessions 1 and 2; ± 12, 24 and 36 arcmin in session 3). B, The depth of the wedges was defined by manipulating disparity in random dot stereograms, which were viewed through red-green anaglyphs attached to prism glasses. C, Blood oxygenation level-dependent signals were acquired from dorsomedial visual cortex. Slice placement is illustrated here on a near mid-sagittal slice in participant 1. D, Signal changes in response to stimulus delivery (stimulus versus rest) for participant 1, showing that activity is localized to the gray-matter. E, Mean percent-signal change for stimulation versus blank periods across all subjects and sessions (N=16). Error bars represent the s.e.m.. F, Mean prediction accuracy for the discrimination of crossed 'near' vs. uncrossed 'far' disparities across early and dorsal visual areas (two-way classification). Chance level (50%) is indicated by the dashed gray line. Error bars depict the s.e.m. across subjects and sessions (N=16). G, Mean prediction accuracy for the discrimination of individual disparity conditions presented within each session (six-way classification). Chance performance (16.7%) is indicated by the dashed gray line. Error bars depict the s.e.m. across subjects and sessions (N=16).

els and mapped these preferences onto flattened representations of the cortex. This produced cortical maps with an apparent organization: contiguous spaces across the cortical surface share similar disparity preferences (Fig. 4.2A,B, top). Importantly, these contiguous areas did not overlap with regions where the mean amplitude of the BOLD signal was low, suggesting that clustering was not a result of macrovascular contributions (Fig. 4.2A,B, bottom). Moreover, considering disparity preferences at different cortical depths suggested consistent preference, which would be expected for hypercolumn-type organization (Fig. 4.2C,D; note that the scale of the cortical depth axis here is expanded relative to the cortical location plane to aid visualization). While these visualizations are useful in illustrating general spatial profiles of responses, the process of mapping and interpolating the data from the (raw) native fMRI data space to a flattened representation of the cortical sheet can introduce over- (or under-) representation of individual datum. In particular, there are frequent one-to-many correspondences between voxels in the functional space and pixels visualized on the flat cortical surface (i.e. oversampling), which can inflate the degree of clustering observed on these flat maps. Therefore, we sought to evaluate preference clustering by examining the disparity response of neighboring voxels in the (native) functional space, thereby ensuring no over- or under- representation of the data.

To quantify disparity preference clustering, we assessed the similarity between the preference of a central target voxel and that of its neighbors. We did this by calculating the distribution of disparity preferences in the population of voxels that shared at least one vertex with the target voxel. Thereby, we calculated a probability map for the disparity preference of the neighborhood referenced to the disparity preference of the target voxel (Fig. 4.3A, for a schematic illustration). Our logical expectation was that if there is clustering in the disparity preferences, target voxels will be surrounded by neighbors with the same- (Fig. 4.3A, top) or similar- (Fig. 4.3A, middle) preferences, in contrast to randomly organized preferences (Fig. 4.3A, bottom). However, the extent to which this structure will be visible depends on the spatial scale of the underlying neural maps in relation to the fMRI sampling resolution. Before examining the empirical data, we therefore consider the extent to which clustering can be recovered based on a simulated data set. To test for clustering at the voxel level, we performed simulations using a model of cortical columns for orientation[266], as there is no standard model for disparity organization. We supposed neural maps of

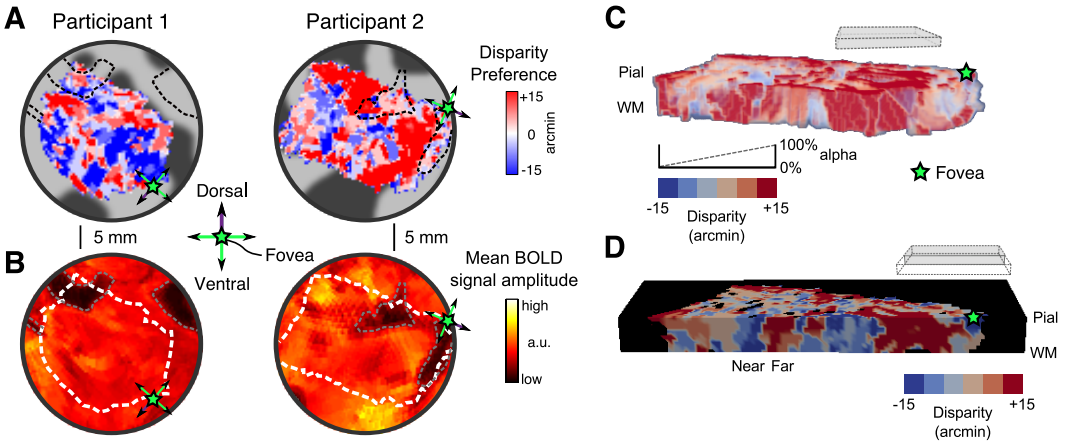


Figure 4.2: Spatial distribution of peak disparity responses in area V3A for two participants. A and B, (top) Peak disparity responses in left V3A of participants 1 and 2 (first session). The peak disparity response of each voxel is mapped onto flattened representations of the cortex. Dark and light gray areas represent sulci and gyri, respectively. Peak disparity responses were sampled from three intermediate layers of the cortical sheet (at relative depths of 0.4, 0.5 and 0.6) and averaged across depths. (bottom) Mean BOLD signal amplitude in the same regions of interest. Dark areas indicate low signal amplitude, and are likely to represent large veins. The white dashed line represents the outline of left V3A shown above. Gray dashed lines delineate areas with low signal amplitude in both maps. Coarse clusters of peak disparity responses do not overlap with the potential location of large veins. C, The same ROI in Participant 2, but now represented across 11 relative points through the entire range of the cortical sheet (0 to 1 relative depth, sampled at increments of 0.1). The flattened representations for each cortical depth were stacked together and an opacity gradient was applied to aid visualization of peak disparity response across the cortical depth. Note that to assist visualization the cortical depth, dimension is not drawn to scale. D, Sliced view of peak disparity responses in the same ROI (Participant 2, left V3A). Data are cut through the cortical depth along a line extending from the foveal representation of V3A up to the periphery near the border with V3d.

different spatial scales (columns from 1 to 4 mm in width) and then sampled these maps using a simulated 1 mm isotropic 'voxel' grid (Fig. 4.3B). Thereafter, we computed the voxel similarity of each sampled voxel relative to its neighbors, and then averaged together the neighborhood preferences of all voxels that had the same central voxel preference. This resulted in a similarity matrix that shows the statistical relationship between the preference of central voxels relative to their surround (Fig. 4.3B, bottom), where strong diagonal structure indicates a close relationship between central voxels and their local neighbors. (Note that the higher probabilities in the top left, and bottom right corners of these plots arise because orientation is a circular dimension; we would not anticipate these for binocular disparity which is a more linear dimension). These simulations indicate that using 1 mm isotropic voxels, it is realistic to obtain information about the structure of underlying cortical organization if the scale of the neural maps is in the region of 3 mm. This corresponds to estimated scale of disparity maps in human cortex based on scaling up measurements from macaque MT to account for overall brain size[100,271]. Having demonstrated proof of concept, we now return to the empirical fMRI data. In principle, we could calculate clustering in exactly the same way as described for the simulations. However, real fMRI voxel responses are not temporally or spatially independent, because of the point-spread function (PSF) of the BOLD signal, meaning that a more sophisticated method is required. In particular, we estimated the preference of the (i) central target voxel and (ii) its neighbors using independent data sub-samples (leave-two-out cross-validation), such that shared preferences for a given measurement could not simply be due to the dependency of BOLD responses for nearby voxels. While this strategy does not remove the influence of spatial blurring, it eliminates temporal correlations between neighboring voxels since we use different time-courses for estimating the preference of central voxels and their surround. We computed preference similarity for each presented disparity value, creating matrices for each region of interest (Fig. 4.3C). We found that diagonal structure in the preference similarity matrices became increasingly apparent for measurements at increasing levels of the dorsal cortical hierarchy. To quantify this observation, we used a reliability statistic that compared the mean probability along the positive diagonal of a matrix, with a distribution of mean values calculated from random sampling from all locations within that matrix (bootstrapping: 10,000 resamples of six values). We found evidence for significant clustering

in V2d (p=.04), V3d (p=.01), V3A (p=.02) and V3B/KO (p=.003), but not in V1 (p=.11). As a control, we re-computed matrices after shuffling disparity preferences and found no systematic structure (Fig. 4.3D). This confirmed that evidence for clustering in higher dorsal areas could not somehow derive from differences in size and/or shape between different regions of interest. Together, these data provide evidence that clustering of disparity preferences is particularly marked in higher dorsal visual areas V3A and V3B/KO in contrast to primary visual cortex. It is possible that preference clustering is much less pronounced in early visual areas. However, recall that from our simulations of maps with different spatial scales, it is possible that there is systematic organization in early areas, but the fMRI sampling resolution does not allow this to be detected.

Testing for the reproducibility of disparity preferences

The preceding analyses support the notion that responses to disparity are clustered in dorsal extra-striate cortex. However, to determine the extent to which this clustering represents genuine cortical structure, we next sought to test whether the spatial distribution of disparity preferences is a persistent property of neuronal responses. Specifically, we tested whether preference maps could be reproduced between different imaging sessions. Comparing functional data across different imaging sessions, especially at very high resolution, is extremely challenging, and previous UHF studies have therefore focused on repeatability within sessions[255,256]. In particular, differences in voxel slab positioning in relation to the cortical sheet affect sampling[255]. Moreover, with a functional resolution of 1 mm (near isotropic), we expect to acquire approximately two points from a given location on the cortical sheet meaning that additional discrepancies could arise from sampling at different cortical depths. As a result, we would not necessarily expect one-to-one voxel correspondence between functional data acquired in different imaging sessions. Nevertheless, we were able to capture similarities between individual disparity preference maps across sessions for four of the six participants that took part in repeated sessions (Fig. 4.4A-D). In these maps, disparity preferences appear to be coarsely organized into bands, which can be clearly identified in maps obtained in different imaging sessions (see the outlines in Fig. 4.4). For one participant (Participant 5, Fig. 4.4E), we found similar structures

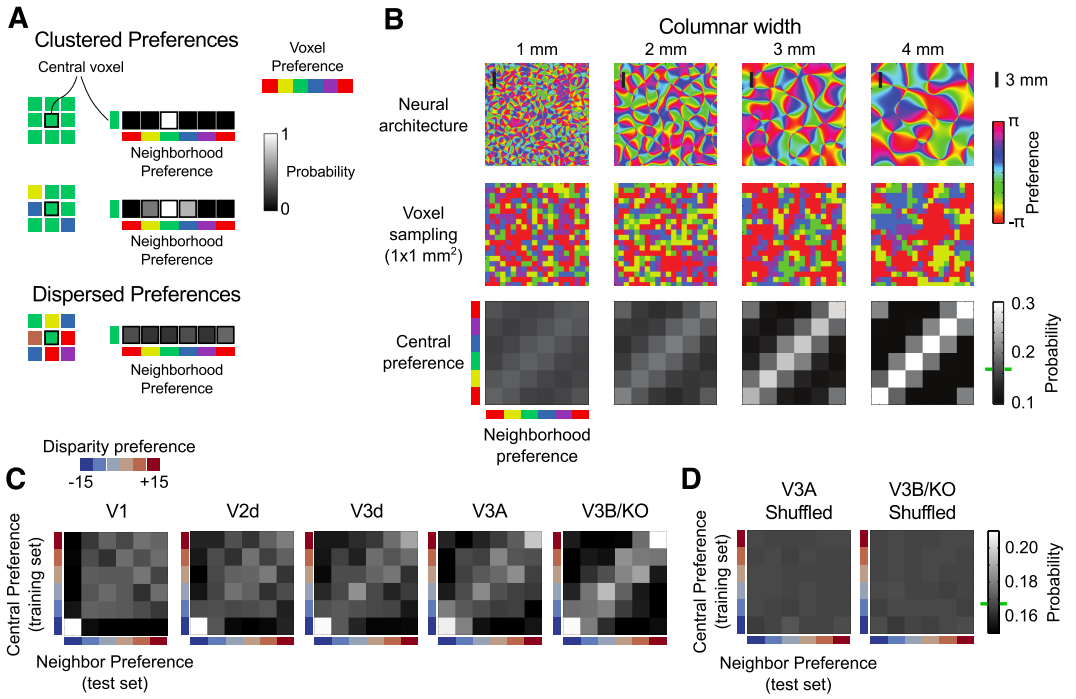


Figure 4.3: Local clustering of peak voxel responses to disparity ('preferences') in simulated and empirical data sets. A, Simplified 2D illustration of clustered and disperse preferences for a given voxel. Individual voxels and their neighbors will often share a similar preference if there is spatial clustering (top and middle). If there is no organization, no relationship should be observed between the peak responses of a target voxel and its neighbors (bottom). B, A simulation of columnar architectures for orientation [266] with periodicity varying from 1 to 4 mm (top) and the respective preference maps after simulating voxel sampling using six equidistant conditions (middle). Bottom: the correspondence between the preference of target voxels and their neighborhood is shown in form of a probability matrix for each columnar width. In each matrix, the i^{th} row represents the average probability distribution of preferences around voxels preferring the i^{th} disparity, and the probability value is represented in grayscale (green horizontal line on the colorbar indicates chance level, 0.167, given that we have considered six preferences). Maps that are well clustered display a clear diagonal structure, demonstrating that nearby voxels tend to share similar preferences. C, The same matrix representation for empirical disparity maps from visual areas V1, V2d, V3d, V3A and V3B/KO. Green horizontal bar on the colorbar indicates chance level (0.167). A diagonal structure emerges along the dorsal cortical hierarchy. Note the different grey scale range from part B for empirical fMRI measurements. D, Results of a similar analysis after randomly shuffling the disparity preferences in V3A and V3B/KO. In this case, we do not observe any diagonal structure.

across sessions, but with reversed disparity sign (note the correspondence between blue and red regions across sessions, particularly in the right hemisphere). This map inversion is consistent with a change from a rear- to front- projection setting, causing a left-right horizontal flip and thereby reversing the disparity sign presented during the experiment. We suspect that this was the result of restarting the projector immediately prior to this participant's scan due to technical problems. For the final participant, we did not find apparent correspondence between sessions, although we note that slice positioning was not optimal in the second session and a portion of V3A was omitted (Fig. 4.4F).

To quantify the similarity between maps, we used voxel-wise correlation in the native functional space. We first selected voxels that had a stable within-session preference and then brought the functional data from each session into common alignment (see Methods and Materials). We then computed the Pearson correlation between corresponding voxels (nearest neighbors) across sessions using bootstrapped resampling (10,000 samples). Confirming our observations from the flattened cortical representations (Fig. 4.4), we observed reliable correlations between disparity maps for four participants (Fig. 4.5A, Participants 1 to 4). In addition, we found reliable negative correlations for one participant (Fig. 4.5A, Participant 5), in line with the apparent inversion of the disparity maps (Fig. 4.4E).

As discussed above, small spatial misalignments between sessions can lead to an underestimation of between-session repeatability. To ameliorate small misalignments, we considered an additional processing stage in which we incorporated a preference-based between-session alignment step. In particular, we calculated an affine transform between the three-dimensional maps that sought to improve co-registration, while minimizing non-linear deformations (see Methods and Materials). We then recomputed correlations across sessions (Fig. 4.5B), and found a small improvement in correlation values for participants with previously reliable between-session correspondence. However, the method itself did not introduce significant correlations (e.g. participant 6) when there was little common structure before alignment, and, in general, the effect of this additional alignment step was quite slight. In particular, Figure 4.6 shows the V3A map for participant 4 (that showed the maximum benefit from this alignment step) with and without the additional alignment step.

To provide an additional measure of reproducibility, we computed the mutual in-

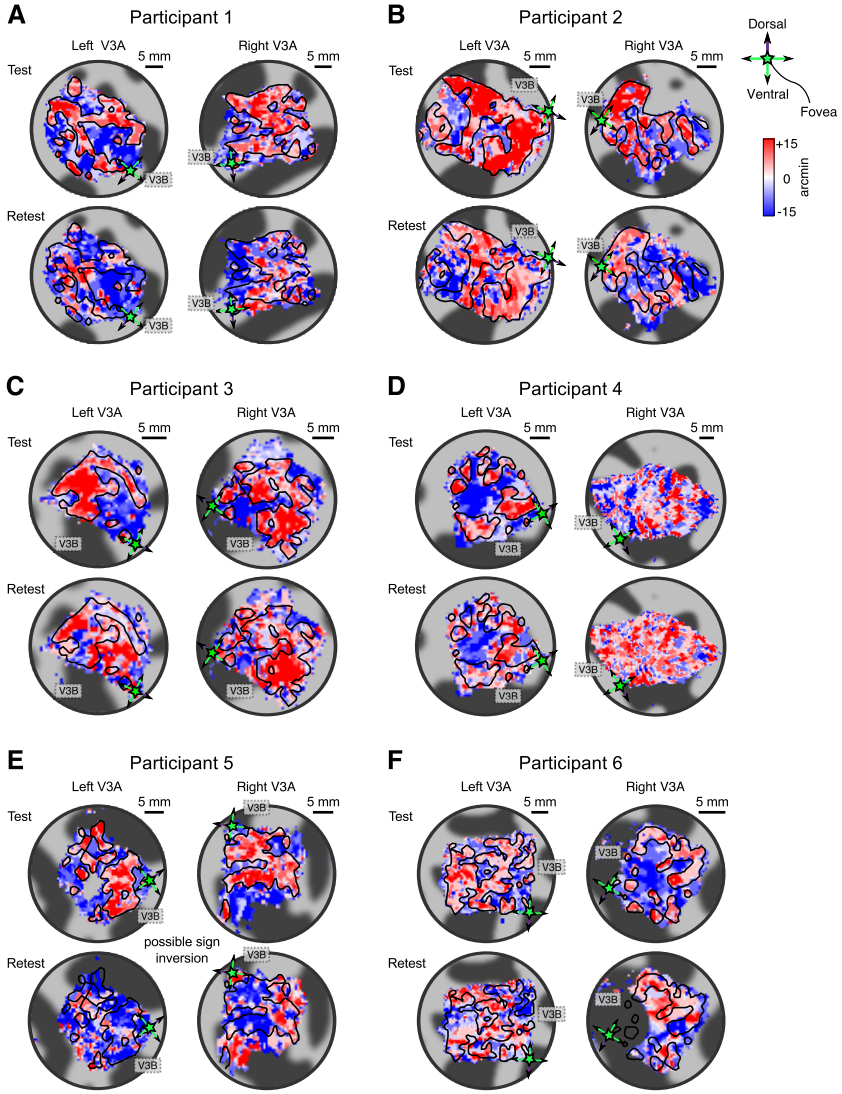


Figure 4.4: Maps of peak disparity responses from area V3A obtained in different imaging sessions. Flattened representations were obtained by averaging disparity preferences across three intermediate layers in the cortex (0.4, 0.5 and 0.6), so as to avoid distortions caused by macrovasculature near the pial surface. Green pentagrams represent the location of the foveal representation used to identify the border between area V3A and V3B/KO (using retinotopic mapping). The dorsal direction is indicated by the purple arrow aligned with the vertex of the pentagram. Additional labels indicate the position of area V3B/KO to aid orientation. A-D, Persistent distribution of disparity responses can be observed in four participants (Participant 1, left V3A; Participant 2, left and right V3A; Participant 3 right V3A; Participant 4 left V3A). Overlaid contours represent the edges of the uncrossed disparity/'far' (red) region from session 1. These were calculated by binarizing the maps, and then applying an edge detection algorithm. The outlines were omitted in panel D (right V3A) since the fine scale changes in this map mean that superimposed contours masked the data and therefore hindered visualization. E, The distribution of peak disparity responses in participant 5 reveals similar structures between sessions, but the color map appears to be inverted. F, No correspondence between disparity maps is evident for participant 6. Slice positioning in the second session was not optimal, with the result that not all of right hemisphere V3A was fully sampled (bottom right).

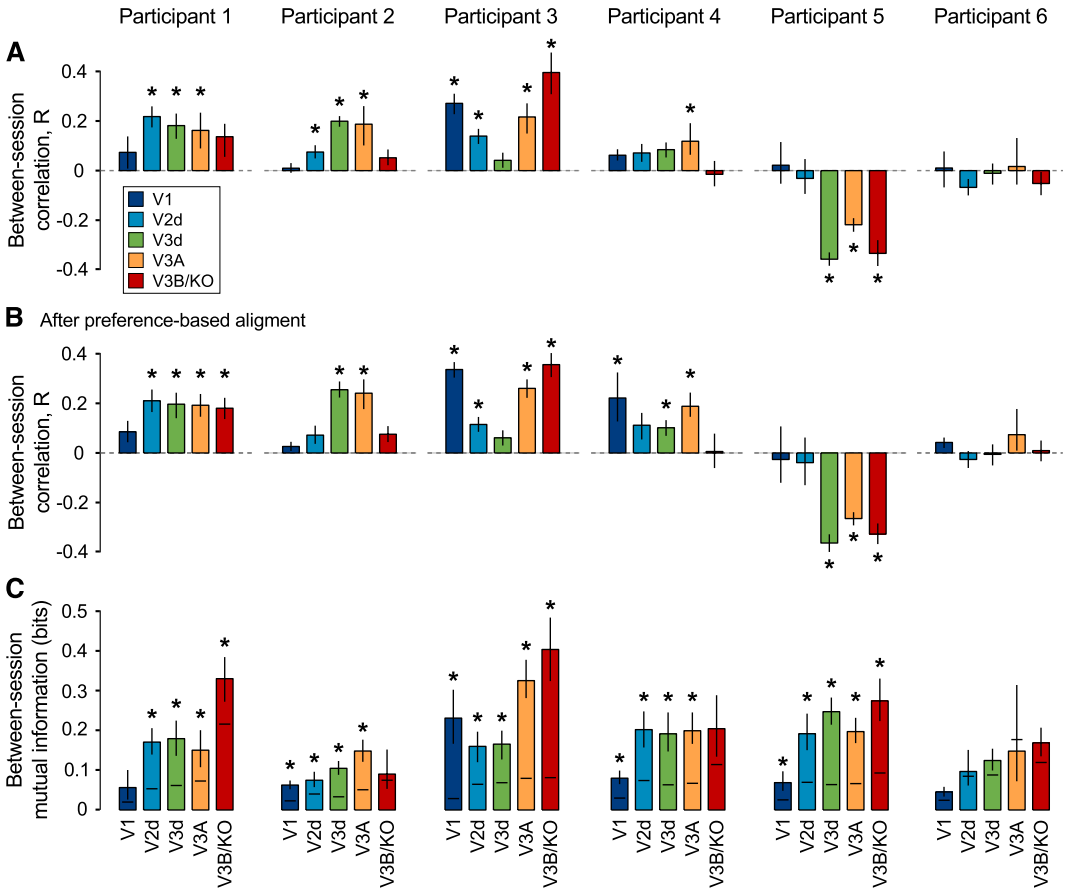


Figure 4.5: Quantifying correspondence between disparity maps acquired in different imaging sessions. A, Correlation between peak disparity responses. Data show the distribution of the Pearson correlation coefficients between voxels' peak responses in session 1 vs. session 2 obtained by bootstrapping (10,000 resamples). The center of the bowties represents the median correlation value. The sides of the bowties and the whiskers extend to the 68% and 95% confidence intervals, respectively. Diagonal slashes indicate that the confidence intervals extend beyond the limits of the ordinate axis. B, As A, except that the alignment between the data in the two sessions was improved using an additional alignment procedure. C, Information shared between maps acquired in different sessions. Bars represent the mutual information between maps, with error bars covering the 95% confidence interval. Horizontal lines represent the 97.5 percentile for a bootstrapped control distribution (10,000 estimates) calculated after randomly shuffling labels of peak disparity response.

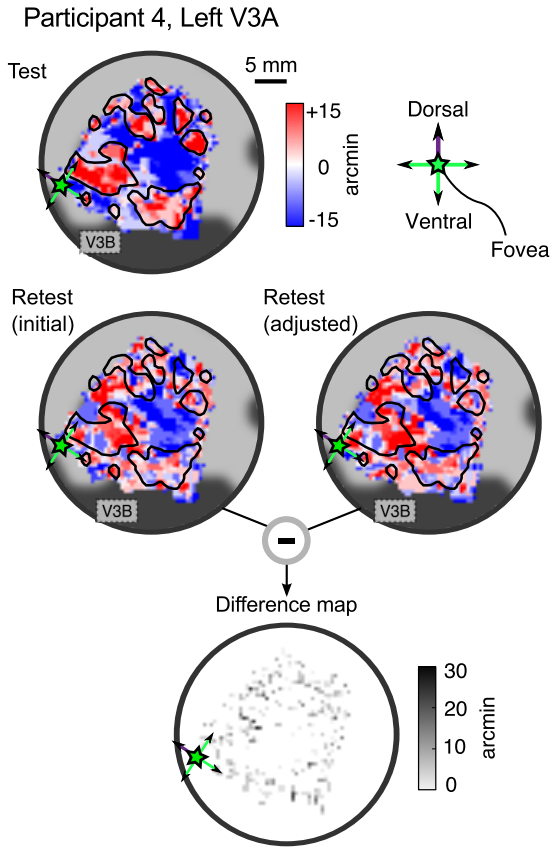


Figure 4.6: Spatial adjustment introduced by the additional alignment step illustrated for participant 3, left V3A, which showed the maximal benefit of this procedure. Top, the map of peak disparity responses obtained in the first imaging session (reproduced from Fig. 4.4D). Middle, maps obtained in the second imaging session before (left, reproduced from Fig. 4.4D) and after (right) adjustment using the additional alignment step. Bottom, a difference map illustrating that only minor differences are introduced by the additional alignment step.

formation[196,268] between maps obtained in different imaging sessions (using the non-preference-aligned data). We compared the empirical mutual information (Fig. 4.5C) with bootstrapped estimates based on randomly permuted disparity preferences (Fig. 4.5C, horizontal lines). We found evidence of persistent information for five participants, confirming the presence of disparity selective structures.

Quantifying voxel response profiles at different disparity magnitudes

In the previous sections, we tested for clustering of disparity preferences by assigning a single disparity preference to each voxel using a winner-takes-all labeling approach. This is an obvious simplification because neurons sensitive to disparity respond to a range of different disparity values. Moreover, disparity tuning curves of individual neurons often vary greatly in morphology — some neurons may respond to a limited range of disparities, while others respond to many horizontal disparities and are only selective for disparity-sign[58,83]. In fact, a comprehensive study of disparity tuning properties suggests that neurons can present a wide range of variation in their responses across the disparity domain. In fMRI measurements, voxels aggregate activity of many such neurons, and consequently the response of individual voxels may present an even greater variety of morphologies. We therefore sought to quantify each voxel's response profile to the range of presented disparities. To do so, we first used the tuning templates described by Poggio[58,83]. These templates offer a descriptive approximation of the response profiles of many disparity selective neurons[73,100], and are simpler (in terms of the number of parameters) than the Gabor models that we will use later. We used these simplified models of disparity selectivity to examine the responses of individual voxels that aggregate the responses of many individual neurons. This voxel-based sampling is quasi-random with respect to the underlying neuronal populations, and, as we discuss above, the scale at which underlying neural representations are sampled clearly influences the information available at the voxel level. Nevertheless, on the basis that disparity representations are clustered, it is reasonable to ask whether the local population activity captured by voxels can be related to physiological models of disparity selectivity, and how such models are distributed across the cortical surface. We considered three types of selectivity (Fig. 4.7A): (i)

near/far tuned responses (Tuned), (ii) near/far categorical responses (Categorical) and (iii) excitatory and inhibitory tuned responses at zero depth (Zero-tuned). For each voxel, we regressed the selectivity models aligned to the preferred disparity of the voxel (Fig. 4.7B) against the beta-weight response profile of the voxel. This provided us with a set of three weights and a constant for each individual voxel. We selected voxels whose activity was well captured by the regression approach ($R^2 > 0.8$), resulting in the selection of approximately half the voxels in each ROI (45 ± 5 %). Of these selected voxels, we found that 65 ± 10 % of the selected voxels were best described as tuned, 15 ± 6 % as categorical and 20 ± 10 % as zero-tuned (mean $\pm$ standard deviation).

Psychophysical investigations and models of stereoacuity suggest that the selectivity of perceptual disparity detectors varies with disparity magnitude — at increased disparity magnitudes, disparity detectors have broader response profiles[84,240]. This led us to hypothesize that such a relationship would be reflected in the representation of 'tuned' and 'categorical' responses at different disparity magnitudes. Particularly, we predicted that increasing the disparity magnitude of our stimuli could lead to an increase in the amount of categorical responses in areas that may be closely related to stereopsis. To test this, we ran an additional experiment with a larger range of disparities. Four participants undertook a third imaging session, during which disparity was varied between 12 and 36 arcmin (crossed and uncrossed). We then used the model-based analysis of the voxel responses to test whether estimated profiles were affected by the increase in disparity magnitude. Specifically, we pooled the data across subjects for each disparity range, and computed the median weight of each model (across voxels). Using this approach, an increased representation of a particular model is demonstrated by an increase in the weight assigned to that model by the regression approach. We visualized the weight for each model in a radar plot with three axes, one for each response model (Fig. 4.7C), and found an increased representation of categorical responses at greater disparity magnitudes, especially in areas V3A and V3B/KO (Fig. 4.7C, compare red versus blue lines; Fig. 4.7D, purple elements). In other words, a greater proportion of the voxels are best explained by the categorical model when the range of presented disparities was larger. This increase in weights for the categorical model was accompanied by small changes in the distribution of tuned and zero-tuned weights (Fig. 4.7D, green and orange elements). Having esti-

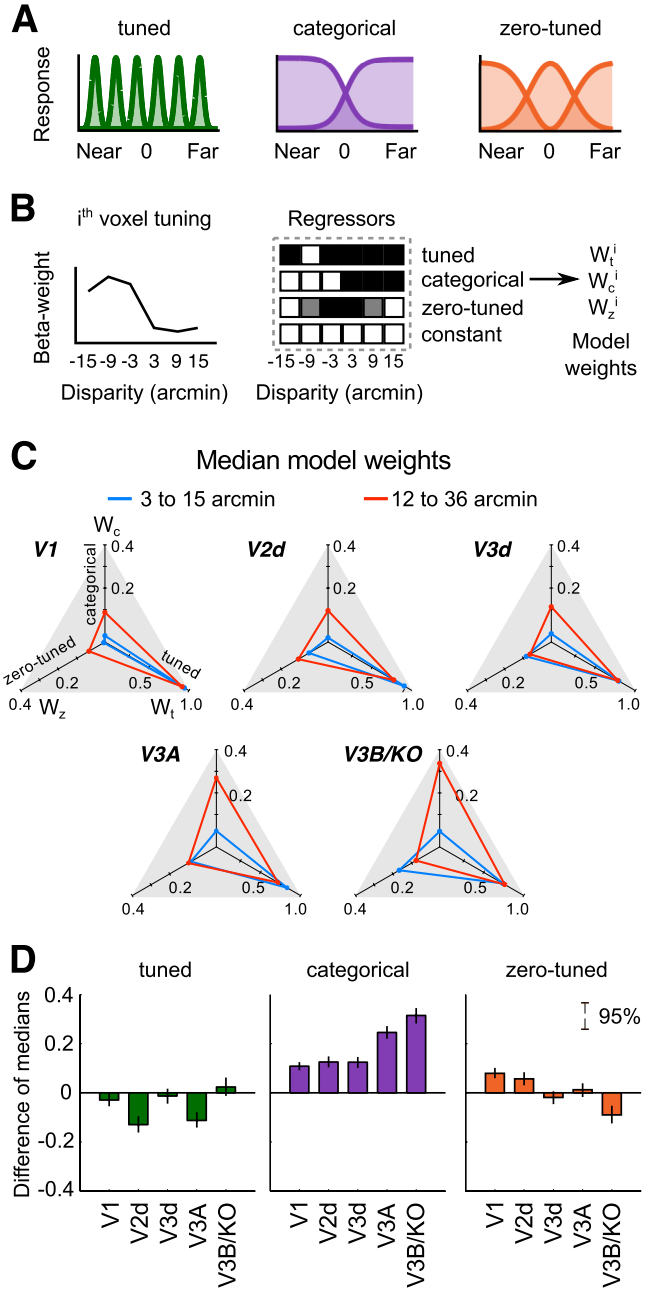


Figure 4.7: Modeling voxel responses using simplified models of disparity selectivity. A, A representation of the descriptive models of disparity selectivity proposed by Poggio et al.[83]. B, Schematic representation of our model-based approach. For each individual voxel, we assembled regressors based on the hypothetical responses for each model type given the peak disparity response of the voxel. After linear regression, we obtain three weights that approximate the contribution of each model for the response profile of individual voxels. C, Model weights at different disparity magnitudes. The median weights (across voxels) for each model are mapped onto a radar plot with three axis (one for each model). Blue lines represent data from the first two imaging sessions (pooled), during which disparity ranged from 3 to 15 arcmin. Red lines represent the distribution of weights observed at disparities ranging from 12 to 36 arcmin (third session). D, Difference in medians between the distributions illustrated in C for all regions of interest. Bars represents the median difference (in medians) obtained by bootstrapping (10,000 resamples). Error bars represent 95% confidence intervals.

mated the type of response profile exhibited by individual voxels, we next sought to determine if there was any structure in the way in which these voxels are distributed across the cortical surface. In particular, we mapped the weights for each selectivity model onto a representation of cortical surface for each individual (Fig. 4.8-4.9). We found three important features in these maps. First, we observed clustering in the weight maps, indicating that nearby voxels share similar disparity response profiles (e.g., voxels described each model are clustered together on the cortex, as shown by co-localized saturated colors). Second, the cortical locations described by categorical vs. tuned models appear to be distinct (Fig. 4.8-4.9, note the complementarity of the green vs. purple maps within-session). Third, the consistency across all sessions was particularly marked for categorical disparity processing model (compare purple maps across sessions one to three, particularly evident in Fig. 4.8A, but also apparent for the other participants), with enhanced categorical representations in session three as expected from the wider disparity range. By contrast, for the tuned and zero-tuned disparity models, we only observed correspondence across the first two sessions in which exactly the same disparity levels were tested (Fig. 4.8B and 4.9A). This highlights the systematic organization of disparity representations, and makes clear that 'tuned' responses are very sensitive to the exact disparity presented, while 'categorical' responses show tolerance to the disparity value.

Our data analysis so far has employed relatively simplistic models of disparity selectivity to describe the responses of individual voxels. Biases in the representation of these models indicated an increase in categorical responses when participants viewed stimuli at higher disparity magnitudes. Next, we sought to examine this relationship in greater detail by fitting physiologically-inspired models of disparity selectivity. In particular, we asked whether changes in voxel response width could be observed between groups of voxels preferring different disparity magnitudes. For each region of interest, we grouped voxels according to their peak disparity response (3, 9, 12, 15, 24 and 36 arcmin, crossed and uncrossed), and then fit a Gabor tuning profile models to each of these twelve groups. Figure 4.10A shows representative responses of voxels with maximal responses for —3, 12, 15 and 36 arcmin, in three different ROIs. For the higher dorsal areas, we observe that the Gabor fit (black line) to the response profile is broader for large disparities than it is for small disparities. We quantified this using the standard deviation (SD) parameter of the Gabor model, plotting this

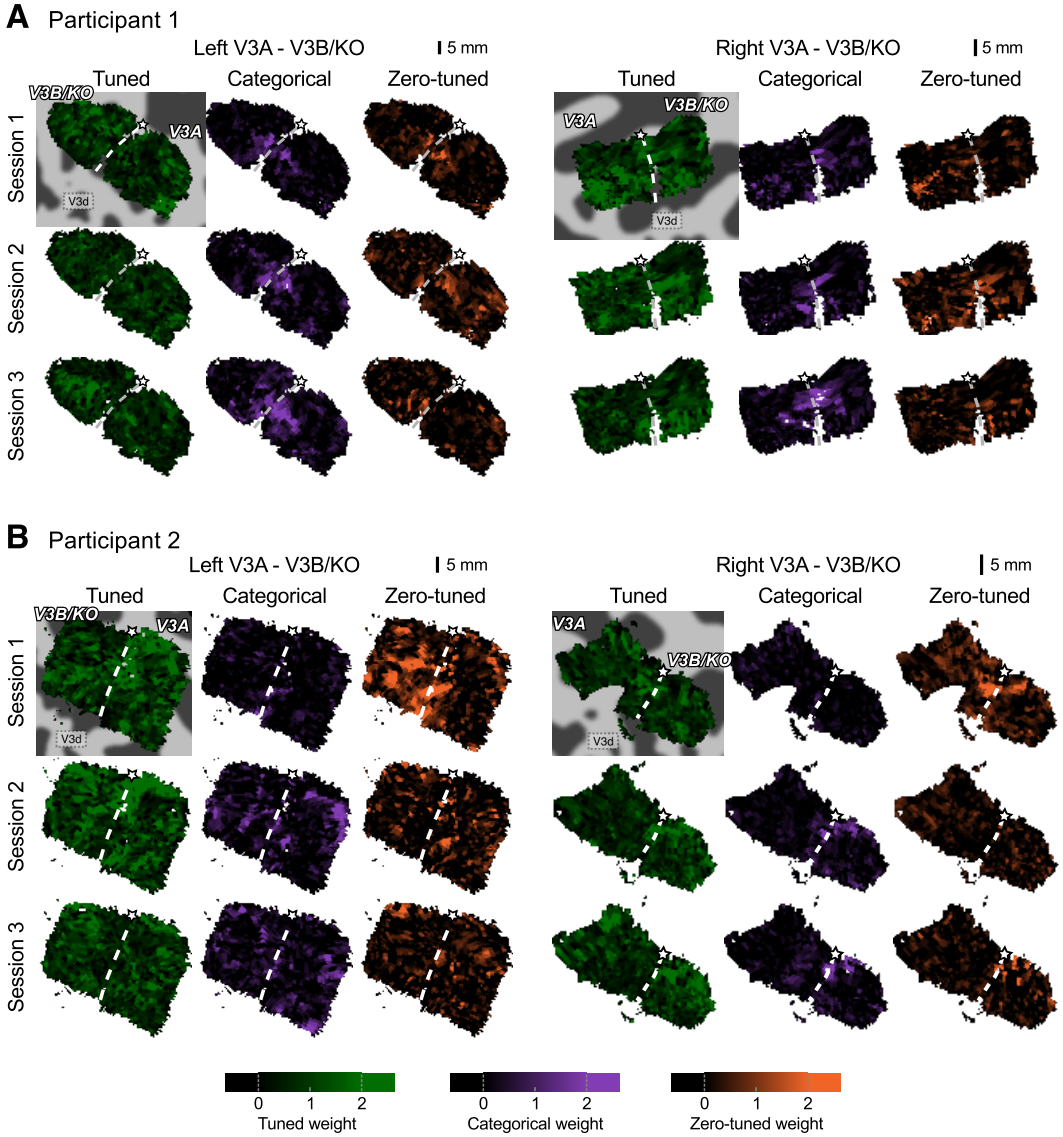


Figure 4.8: Cortical representation of models weights in areas V3A and V3B/KO in the left and right hemispheres of participant 1 (A) and 2 (B). The pentagram on each map represents the position of the fovea, and the white dashed line the division between V3A and V3B/KO established using retintopic mapping. A, Categorical responses (purple) were persistently identified around the foveal representation dividing V3A and V3B/KO (both hemispheres), even when different disparity levels were presented (session 3). B, An apparent correspondence between tuned responses was found across sessions 1 and 2, but not session 3 (left hemisphere).

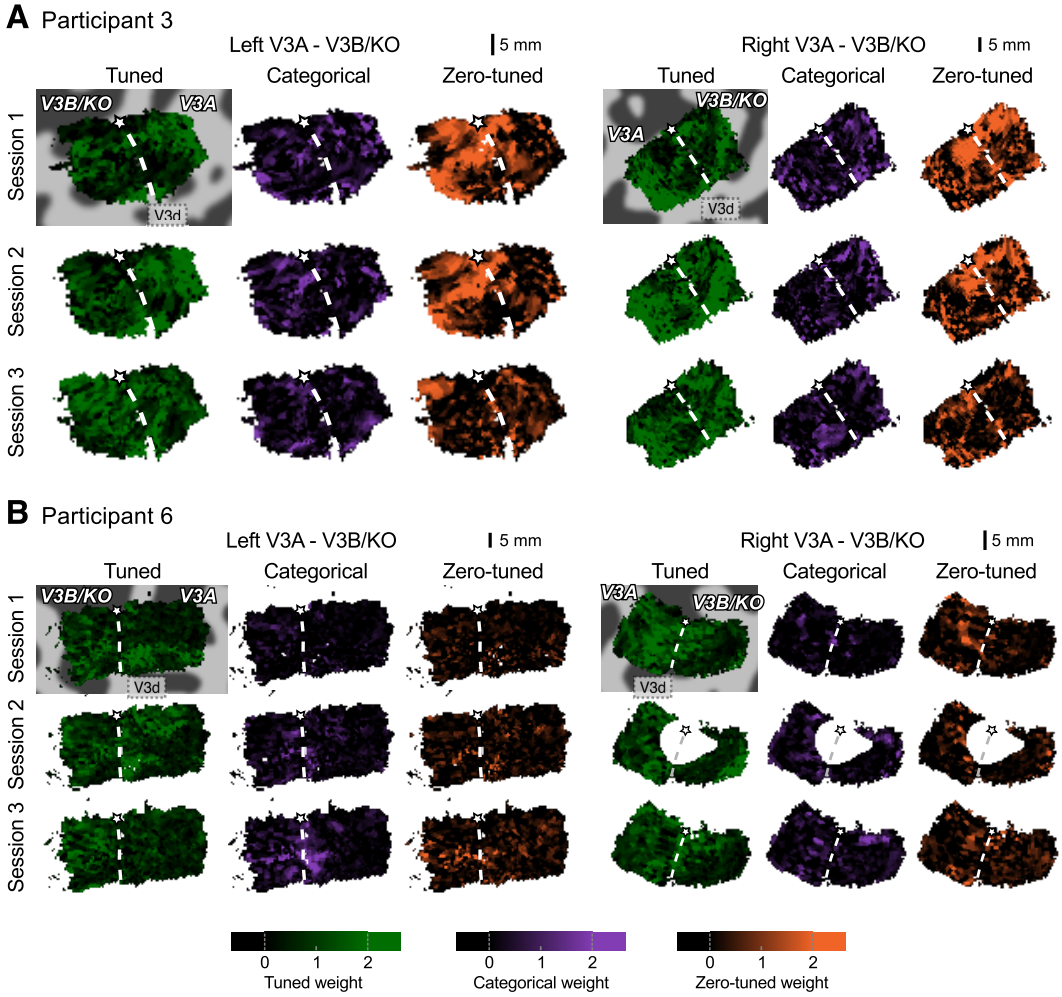


Figure 4.9: Cortical representation of models weights in areas V3A and V3B/KO in the left and right hemispheres of participant 3 (A) and 6 (B). This figure follows the format presented in Figure 8. A, Evident correspondence between 'tuned' and 'zero-tuned' weights was identified across the first two imaging sessions. That correspondence is not observed for the third session, where disparity magnitude was increased. B, Apparent correspondence was absent for participant 6. Note that the voxel slice placement for session 2 meant that we omitted coverage of a considerable portion of V3A and V3B/KO in the right hemisphere.

as a function of the peak response of the voxels (Fig. 4.10B). In early visual areas (V1, V2, V3d) we found that the width of voxels' response profiles was not related to the overall disparity magnitude. By contrast, in areas V3A and V3B/KO we observed significant relationships between the peak disparity response and the width of the Gabor fit. One potential concern with this analysis is that such a relationship may be a consequence of the differential spacing between the presented disparities in different imaging sessions, i.e. the envelop width is broader because of the wider stimulus spacing. Ideally, we would have used a fine spacing of disparities across a broad range of disparity magnitudes, which would rule out such a concern. However, the practicalities of obtaining a sufficient number of fMRI measurements within a time-constrained imaging session meant we could not do this. Nevertheless, we judge it unlikely that stimulus spacing per se accounts for the relationship we observe. First, we did not observe significant correlations in V1 to V3d, suggesting that the increase in disparity spacing alone does not result in a relationship between the peak and SD parameters. Second, we can contrast profiles for overlapping points in this space (Fig. 4.10a): the width of the fit to the −15 tuned units (from sessions 1 and 2 with 6 arcmin stimulus spacing) is wider than for −12 (from session 3 with 12 arcmin stimulus spacing).

Changes in selectivity as a function of disparity magnitude are thought to be characteristic of neural populations that underlie human stereoscopic judgments[84,240] (Fig. 4.11A). Based on our fMRI measurements, we sought to test how well estimates of human neural population responses to disparity could account for depth discrimination thresholds. To this end, we built a population of disparity-tuned units based on the estimated (linear) relationship between Gabor parameters and disparity magnitude (Fig. 4.11B). Using these values suggested that V1 responses were unlikely to account for disparity discrimination judgments (Fig. 4.11B), however, estimated populations in V3A and V3B/KO produce discrimination threshold curves that are qualitatively similar to previously reported behavioral results[239] (Fig. 4.11C,D) and stereo acuity modeling[84] (Fig. 4.11A). While the overall shape of the curves are similar, a closer fit would likely require testing a wider range of disparities to account for the flanks of the curves, and denser sampling near the fixation point to capture the fine trough near zero disparity. Stevenson et al.[240] use psychophysical measurements to describe the relationship between tuning width of the perceptual mechanisms (parameterized

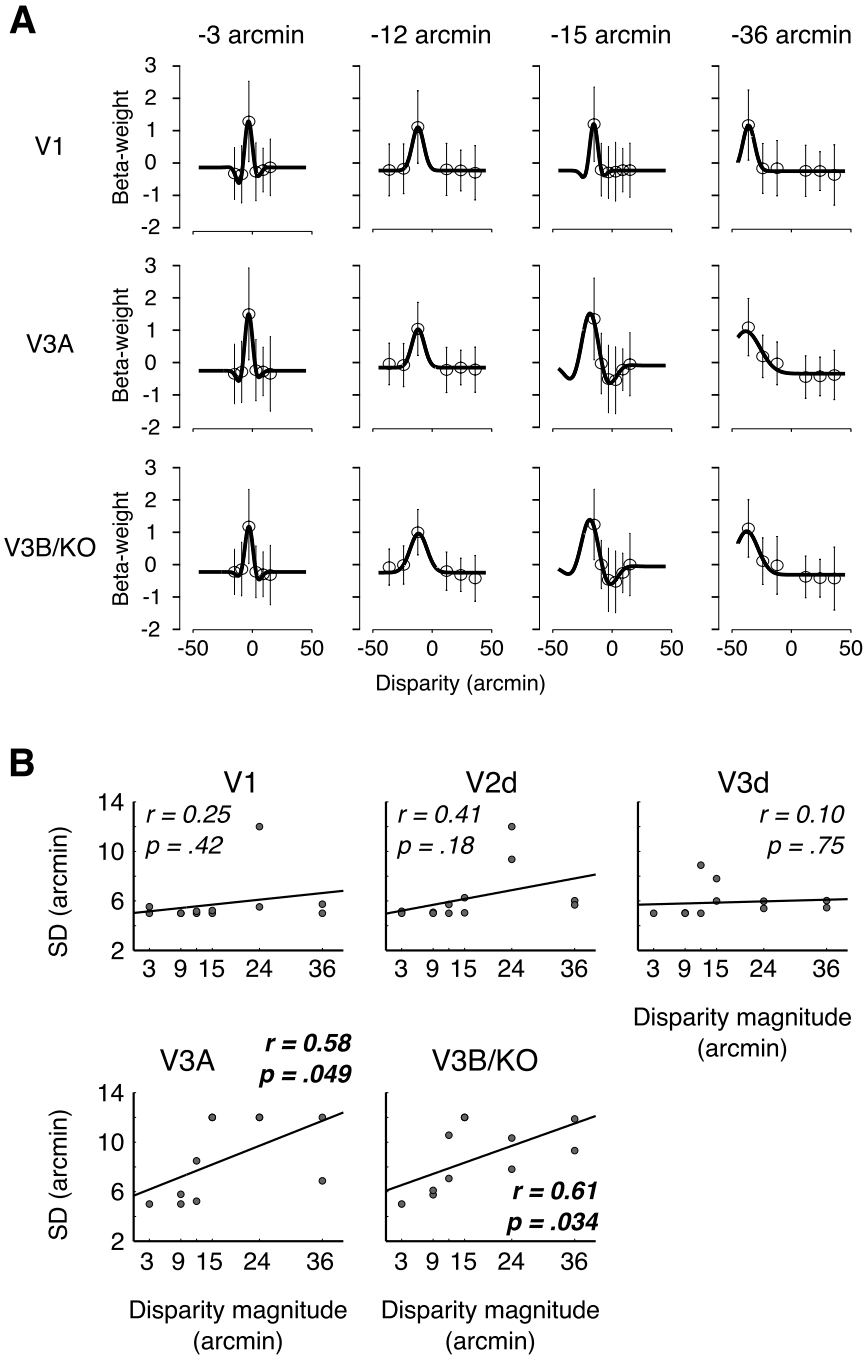


Figure 4.10: Voxel response profiles at different disparity magnitudes. We modeled voxel responses using Gabor filters, and examined the relationship between the Gabor parameters and preferred disparity magnitude. A, Pooled voxel responses in areas V1, V3A and V3B/KO modeled by Gabor filters for four preferred disparities (−3, −12, −15 and −36 arcmin). Gabor models were fit to sets of voxels sharing the same preferred disparity, resulting in twelve groups per ROI. Error bars represent the standard deviation across voxels. B, Relationship between response profile width (standard deviation of the Gaussian envelope) and peak disparity response for early and dorsal visual areas. Each datum represents a group of individual voxels that share the same disparity preference (one of the twelve preferred disparities examined in our experiments: ± 3, 9, 12, 15, 24 and 36 arcmin). A significant positive trend between tuning width and disparity magnitude was found in V3A and V3B/KO, but not in earlier visual areas.

as the full width at half maximum, FWHM) and disparity magnitude. Using data extracted from their paper, and the linear relationship they estimated, we plotted our fMRI estimates of voxel response width (FWHM) together with their data and estimated linear relationship (Fig. 4.11E). This suggests a striking similarity between perceptual- and fMRI- estimates of variations in the tuning of units that respond to binocular disparity. Together, these results suggest an intriguing analogy between activity in V3A and V3B/KO and depth judgments (consistent with previous neuroimaging studies[193,271–273]).

4.4 Discussion

Here we use 7 T fMRI to test whether human dorsomedial visual cortex contains systematic organized representations of binocular disparity. Using a series of computational modeling approaches, we report three main advances in understanding disparity organization in the human brain. First, we show that disparity preferences are systematically organized (Figs. 4.3C, 4.4, 4.8, 4.9), and importantly that these preferences are persistent between imaging sessions (Figs. 4.4, 4.5, 4.8, 4.9). Second, we observed differences between the local distribution of disparity responses in early and dorsomedial visual areas (Figs. 4.3C, 4.7, 4.10), suggesting different properties of cortical organization. Third, by modeling the responses of individual voxels, we show a relationship between tuning width and disparity magnitude (Fig. 4.10B), indicating more broadly tuned responses to larger disparities, in line with psychophysical and modeling work that posits such a relationship as a characteristic property of neural populations involved in stereopsis (Fig. 4.11). Together, these findings indicate that human V3A and V3B/KO contain selective cortical structures that are likely to be important in stereoscopic depth processing. The cortical organization of disparity preferences Understanding of the cortical structures that support disparity processing is largely informed by recordings in the macaque brain. For instance, by systematically assessing disparity preferences at locations across the cortical surface of area MT/V5, DeAngelis and Newsome[100] demonstrated smooth changes in preferred disparity across the cortex that indicate systematic organization. More recent electrophysiological evidence indicates that other dorsal visual areas, including V3A, contain clustered representations of disparity[98,149]. Based on previous human

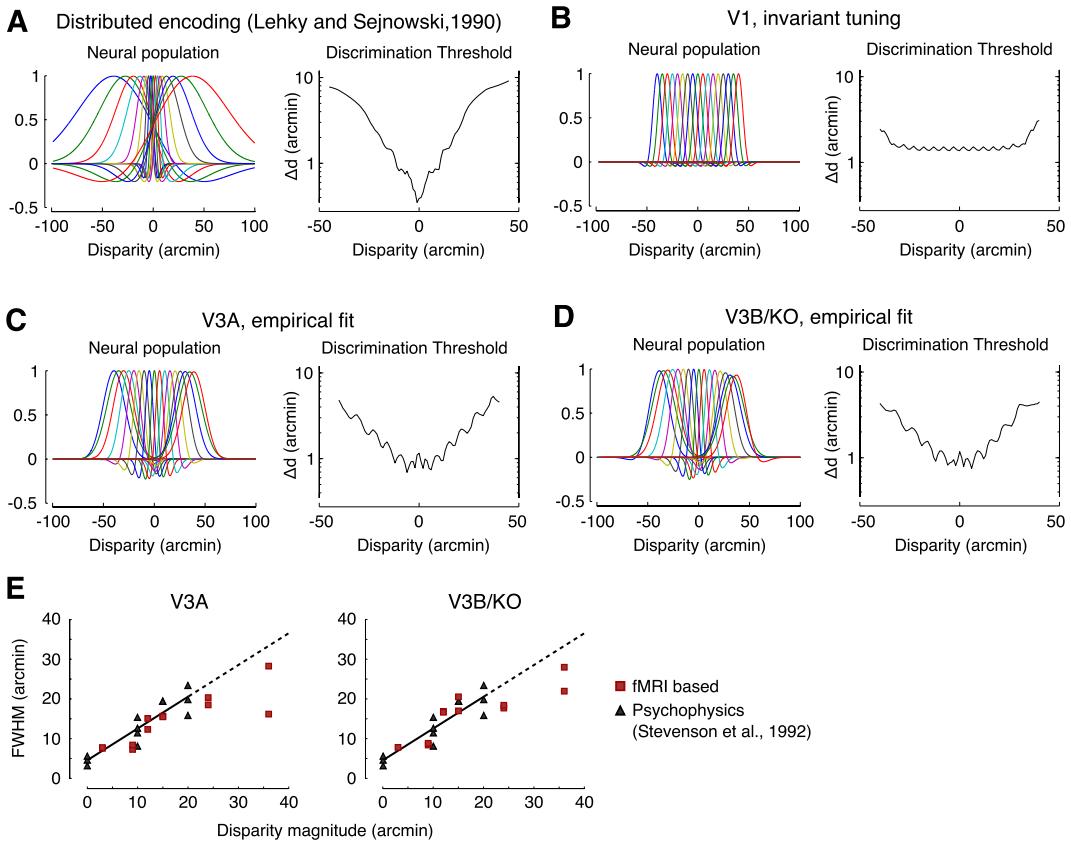


Figure 4.11: Population encoding mechanisms and stereo acuity at different disparities. A, Distributed encoding model proposed by Lehky and Sejnowski[84]. A population of seventeen non-uniform, largely overlapping units (left) produces a disparity discrimination curve (right) similar to stereoacuity judgments made by human observers[239]. B, Interval encoding model derived from voxel response profiles in V1. A population of seventeen units with uniform, narrow tuning produces a disparity discrimination curve uncharacteristic of the human visual system. C, D, A neural encoding model derived from the voxel response profiles in areas V3A and V3B/KO (left), and the simulated discriminative performance of these models (right). The performance of these models is more similar to the idealized patterns of psychophysical performance (part A) than a model derived from V1 activity (part B). E, Plot of disparity magnitude against detector tuning width based on psychophysical data published by Stevenson et al.[240], and our fMRI estimates in V3A and V3B/KO. The trend line reproduces that fit by Stevenson and colleagues, with black data points representing their published data (their Figure 7) as obtained by a 'data thief' procedure implemented in Matlab. Red data points represent fits from on our fMRI measurements. The dashed portion of the fit extends the line of best fit beyond the range of disparities tested by Stevenson et al.[240]

imaging work[129,193], dorsomedial visual cortex was likely to show strong responses to disparity-defined stimuli in human participants. We find evidence that similar disparities are clustered together, particularly in areas V3A and V3B/KO (Fig. 4.3C), and that correspondence between maps can be observed even when the presented disparities differ (Fig. 4.8-4.9, categorical responses). While we have concentrated on area V3A, it is interesting that our findings suggest cortical organization for disparity that is similar in its basic properties to macaque MT. Although these two areas appear to have distinct functional properties for binocular disparity[274], they receive a large portion of inputs from common areas[50]. In particular, MT and V3A receive inputs from V2 and V3, where disparity organization has previously been reported[98,146,148,252]. Therefore, it is possible that cortical organization for binocular disparity in dorsomedial areas and MT is derived from their downstream inputs in the cortical hierarchy. While our imaging data suggests clustered responses, it is clearly not possible for us to infer that the underling organization is columnar. For instance, it is possible that the persistent structures we observe across sessions represent a coarser spatial bias in disparity responses, rather than a periodic columnar structure. Nevertheless, it is encouraging that we observed a difference between dorsal visual areas and responses in primary visual cortex. This appears consistent with macaque electrophysiology in suggesting V1 has only a weak tendency for clustering[143,144].

Benefits and limitations of UHF imaging for mesoscopic mapping

Our understanding of cortical organization to date is predominantly informed by animal models, typically using neurophysiological and optical imaging methods that provide a high level of detail, at cost of invasiveness. Recent advances in ultra-high field fMRI make it possible to investigate mesoscopic properties of the human cortex non-invasively[255–257]. However, issues in the interpretation of neuroimaging data are usually introduced by (i) potential biases from large vascular structures, which are poorly related to local cortical activity and (ii) insufficient spatial resolution. Here, UHF fMRI revealed that disparity preference representations are well clustered in human dorsal visual cortex. While we may be able to map disparity preferences at lower spatial resolution, the increased BOLD CNR of UHF is likely a requirement to do so. We have previously attempted to investigate disparity maps at 3T,

but without success (unpublished observations). By imaging at UHF, the contributions from large vessels, which could mask the actual distribution of disparity preference, are reduced[260–262]. This gain in spatial specificity is the fundamental benefit for mapping genuine properties of neural subpopulations. Although our UHF imaging improves spatial specificity, additional care is necessary to avoid the influence of large vessels, especially when mapping functional data onto cortical flattened maps: mapping large veins onto flat maps can result in the emergence of spurious structures unrelated to local activity. We therefore chose to sample functional activity predominantly from the central layers of the cortex in order to avoid large surface vessels[257,263], and improve spatial localization[264]. This is particularly important given that we used a gradient-echo sequence, which is more susceptible to surface macrovascular contributions compared to spin-echo based sequences[275]. Additionally, we verified that regions where the mean BOLD amplitude was higher (which could derive from larger vessels) were not co-localized with coarser structures found in preference maps (Fig. 4.2A,B). Finally, it is necessary to consider the possibility that clustering is enhanced by the point-spread-function (PSF) of the 3D GE-EPI sequence. The PSF of the BOLD signal can reach 2 mm in extent (Gaussian FWHM)[276], meaning that voxel responses may be significantly influenced by the activity of their neighbors. However, it is unlikely that the PSF is a major barrier to the interpretation of our data. First, even a sequence with a broad PSF can be used to map cortical properties, provided that the contrast-to-noise (CNR) is sufficient[256]. Second, limiting our analyses to voxels from central layers of the cortex (i.e., away from large draining vessels on the cortical surface) is likely to have reduced the spatial spread of the BOLD response[264]. Finally, our data point to differences in disparity clustering between visual areas (Fig. 4.3c). This suggests that our measurement approach has sufficient dynamic range that we can capture changes related to the underlying structure of the cortical organization.

Disparity selectivity and stereopsis

Models of human stereo acuity have posited a relationship between the tuning width of disparity sensitive units as a function of the magnitude of disparity: i.e., neurons selective for fine disparities have smaller receptive fields, while units preferring coarser

disparities have larger receptive fields (Fig. 4.11A;[84]). Psychophysical measurements support this conclusion[240]. In our study, we found that the population-estimated responses in human V3A and V3B/KO follow this relationship, and a model based on fMRI estimated tuning widths as a function of presented disparity is able to discriminate disparities in a manner similar to the human visual system. This is captured by the slope of the disparity discrimination curves between small and large disparity magnitudes for V3A and V3B/KO (Fig. 4.11C,D), resulting in greater stereo acuity for fine rather than coarse disparities. Conversely, a population with invariant tuning properties produces nearly constant disparity discrimination thresholds, implying constant stereo acuity for a wide range of disparity magnitudes (Fig. 4.11B) In order to examine disparity responses, we used well-defined tuning templates to group voxels according to their response type (e.g. tuned vs categorical). These templates can be seen as simplifications of the tuning classes suggested by Poggio and colleagues more than two decades ago[83]. Since then, it has been suggested that disparity selectivity is better described by Gabor models whose (continuous) parameter space explains previously posited discrete types of disparity tuning[73]. When we used Gabor models to describe the voxel responses for each disparity level, we found that changes in the envelope width along the disparity domain can be well approximated by a linear function (Fig. 4.11E; consistent with psychophysical investigations[240]), suggesting that tuning width varies gradually with disparity magnitude (at least within the range we have tested).

4.5 Conclusion

Using 7 T fMRI, we show that human dorsal visual areas contain systematically organized structures for disparity processing. The responses of these structures vary with disparity magnitude, which aligns well with previous quantifications of stereoscopic perceptual judgments. Together, our results suggest that areas V3A and V3B/KO contain selective, organized structures that support stereoscopic processing in the human brain.

Chapter 5

Layer-dependent Activity in V1 during Stereopsis

5.1 Introduction

The positional difference between the images captured by the left and right eyes, known as binocular disparity, is a highly informative cue to depth perception[12,150]. The process of estimating depth from binocular disparity (i.e. stereopsis) is a complex process that requires multiple computational steps[158]. Many different cortical areas have been implicated in stereopsis[95,136], but their precise roles remain unclear.

Disparity processing is thought to begin in primary visual cortex, where many disparity selective neurons have been found[45–47,58]. Importantly, simple subunit models are able to explain disparity selectivity in V1 relatively well, and have been instrumental in building theories of early disparity processing[63]. However, the activity of single neurons in V1 is poorly related to stereoscopic perception[87,277] — in contrast with neurons in many extrastriate areas, such as V2[117,118], V4[102], MT[278] and dorsomedial visual cortex[51,129,193]. To date, we have little data to elucidate how striate and extrastriate areas interact to support stereoscopic perception.

One promising avenue towards understanding the mechanisms by which visual areas interact is characterizing neural activity at the level of cortical layers. Because different cortical layers have largely dissociable patterns of feedforward, lateral and feedback connections, mapping neural activity at different cortical layers can reveal the

flow of information (e.g. top-down or bottom-up) associated with a particular stimulus or task. For instance, recordings performed with laminar probes in macaques have helped developing new theories of how early and higher visual areas support object-based attention[279] and working memory[280]. To our knowledge, this technique has not been yet used to characterize the laminar profile of neural activity evoked by stereoscopic stimuli.

Invasive and localized recording techniques require *a priori* specification of a limited number of sampling sites, which hinders the investigation of multiple visual areas simultaneously. This is particularly important for investigating stereopsis given the multitude of cortical areas that have been implicated[193]. Ultra-high field functional resonance imaging is emerging as a promising technique to investigate the role of different cortical layers over extended regions of cortex. For instance, layer-dependent fMRI has been successfully used to extract layer-dependent BOLD signals in response to illusory figures[281] or partially occluded scenes[282].

Here, we set out to test the feasibility of using ultra-high field imaging at submillimetre resolution to investigate layer-dependent BOLD signals associated with stereoscopic perception. We compared the response to correlated random-dot stereograms, which elicit perception of a surface in depth, and contrast this against responses to anticorrelated random-dot stereograms, for which a surface in depth could not be perceived. We found that BOLD signals in extrastriate cortex were a strong correlate of stereoscopic perception. Conversely, little differential BOLD signals were observed in V1. Layer-dependent sampling revealed a bias towards superficial layers in LO, while an effect of cortical depth was not consistently observed in V1. Our results suggest that cortical layers in V1 and LO are differentially recruited during stereoscopic viewing and — more generally — that ultra-high field imaging at submillimetre resolution is a promising technique to investigate the role of different cortical layers in stereopsis.

5.2 Materials and Methods

Participants

Eight subjects aged between 25 and 40 years (five male) participated in the study. Participants provided informed consent and procedures were approved by the Ethics Committee of the Faculty of Psychology and Neuroscience at Maastricht University. All participants had normal or corrected-to-normal vision and did not present stereo deficits.

Stimuli and design

Stimuli were presented using a fiber optics goggle system (Silent Vision SV-7021, Avotec, Inc., Stuart, FL). The resulting visual field was approximately 30° horizontally by 23° vertically, and the viewing distance was approximately 6 cm. Stimuli consisted of random-dot stereograms (12° x 12°, 34 dots/deg^2) with a mid-gray background. To promote stable vergence, the stimuli were surrounded by a static grid of squares. Dots in the stereogram followed a black or white Gaussian luminance profile, subtending 0.07° at half maximum. In the center of the stereogram, four wedges were equally distributed around a circular aperture (1.2°), each subtending 10° in the radial direction and 70° in polar angle, with a 20° gap between wedges (Fig. 5.1). The wedges were presented in correlated or anticorrelated forms and had crossed or uncrossed disparity (10 arcmin, $\pm$ 0.5 arcmin jitter). The surrounding was always presented in correlated form at zero disparity. At a given time point, all wedges were rendered with the same disparity and binocular correlation. To reduce adaptation, we applied a random polar rotation to the set of wedges such that the disparity edges of the stimuli were in different locations for each stimulus presentation. In the center of the wedge field, we presented a fixation square (side length = 1°) paired with horizontal and vertical nonius lines. Stereo correspondence and disparity sign were held constant during 12 second blocks, during which we presented 10 stimuli (900 ms on, ISI 300 ms). During each acquisition run, we presented each combination of correspondence and disparity sign 10 times, resulting in 40 blocks. In addition, there was a fixation block at the start and the end of each run. Each run lasted 504 seconds (42 blocks x 12 s), and we collected six to seven runs in each imaging session. On each

run, we asked participants to fixate in the central fixation square while performing a Vernier detection task[193]. For localizing activity to stimulus delivery and for quality control purposes, participants undertook one experimental run during which we presented radial checkerboard flickering at 8 Hz. The checkerboard was positioned in the center of the screen and subtended 12 degrees. Participants were instructed to fixate in the center of the screen passively while viewing the stimuli. The checkerboard was presented during 2 seconds, and was followed by a blank period of 14 seconds. A 16 second blank period was included at the beginning and at the end of the run.

Imaging

Imaging sessions were performed at the Maastricht Brain Imaging Center (Maastricht, The Netherlands). We used a 7 Tesla Siemens scanner with a 16-channel surface coil system. Motion was restricted by the use of foam padding. Anatomical volumes were acquired in the beginning of each session (3D-MPRAGE, 256 slices, FOV = 230x230 mm^2, matrix size = 384x384, 0.6 mm isotropic resolution), followed by a proton-density weighted volume with the same resolution and matrix size, which we used for inhomogeneity correction. We then acquired blood oxygen level-dependent (BOLD) signals during stimuli delivery. For the localizer experiment, we used two dimensional echo-planar imaging (gradient-echo; TE/TR = 24/2000 ms; flip angle = 70°; FOV = 150x150 mm^2, matrix size = 136x136; 28 slices; 1.1 mm^3 isotropic resolution). For the main experiment, we acquired BOLD signals using three-dimensional gradient recalled spin-echo imaging with inner volume selection (3D GRASE; TE/TR = 38/2000 ms; FOV = 24x128x9.6 mm^3, matrix size = 30x160x12, APxRLxFH; 0.8 mm^3 isotropic resolution). Slice positioning was guided by visual identification of the calcarine sulcus. The acquisition volume was placed along the calcarine sulcus and efforts were made to cover both banks of the sulcus in both hemispheres, when possible.

Imaging data were analyzed using BrainVoyager QX 2.8.2 (Brain Innovation, Maastricht, The Netherlands) and custom Matlab code (The Mathworks Inc, Natick, MA). Field inhomogeneity in anatomical scans was corrected by dividing the structural volume by the respective proton-density weighted volume. When low frequency variations remained, an additional automatic inhomogeneity correction step was per-

formed. Anatomical volumes were then manually segmented to ensure high-quality cortical representations. The resulting segmented volume was used to compute estimates of cortical thickness following the Laplace method[283], from which we could later derive streamlines at different cortical depths for laminar analysis. Details on this procedure are described elsewhere[257]. Functional data were preprocessed in functional native space to avoid unnecessary data interpolation. Head motion was estimated using trilinear interpolation and subsequently corrected using sinc interpolation. Low-frequency fluctuations were removed using a GLM with Fourier basis set at 2 cycles per run. Preprocessed data were then coregistered to the anatomical space and resampled using trilinear interpolation.

We started our analysis by defining regions of interest that were well driven by stimulus delivery. To do so, we fit a general linear model to the data, and tested for positive differences between BOLD signals during stimulus delivery and rest periods. In particular, we ensured that the region of interest comprised areas that were well modeled during the main experiment (F-contrast, stimulus versus blank, $p<.005$, FDR-corrected) or during the checkerboard localiser run (F-contrast, stimulus versus blank, $p<0.001$, FDR-corrected). Using the cortical thickness measurements in these selected regions, we then computed nine equally spaced grids along the cortical sheet (0.1 to 0.9 in relative depth, with 0.1 increments)[257]. Each grid intersects the cortical sheet at a different depth, and therefore runs approximately along the cortical layers. Next, the nine equally spaced grids were grouped in three equally space bins for deep, middle and superficial layers, and the corresponding cortical depth label was assigned to each voxel according to the nearest located cortical grid.

General Linear Modeling

For the stereo correspondence experiment, we modeled the BOLD signal using four conditions of interest (correlated, negative disparity; correlated, positive disparity; anticorrelated, negative disparity; anticorrelated, positive disparity). Advantage for stereo correlated stimuli was then tested by contrasting activity elicited by correlated versus anticorrelated blocks. These univariate procedures yield a statistic for each voxel, which we can then index by the corresponding cortical depth label assigned above.

Multivoxel pattern analysis

For each region-of-interest, we converted voxel time series to z-scores, and shifted the respective time course by two volumes (equivalent to four seconds) to account for the hemodynamic delay. Then, we took the median value (per voxel) across time within each condition block, resulting in 40 activity patterns per run. We trained a support vector machine classifier with leave-one-out cross-validation (i.e. leaving one run out in each fold). These resulted in 240 patterns for training and 40 patterns for testing the model (less 40 training patterns for one subject for which we acquired only six runs of the main experiment). For each cross-validation fold, we stored the absolute weights assigned to each voxel by the classifier, and examine their distribution at different cortical depths. As mentioned above, we divided voxels in three groups according to their relative distance from the white matter: deep layers (from 0.1 to 0.3); intermediate layers (relative depths from 0.4-0.6); and superficial layers (from 0.7-0.9).

5.3 Results

We recorded blood oxygenation level-dependent signals while participants fixated on a central cross and viewed correlated and anticorrelated random-dot stereograms (Fig. 5.1). The stereograms consisted of four concentric wedges rendered in correlated (Fig. 5.1A) or anticorrelated form (Fig. 5.1B); in the former, the wedges were perceived in depth; in the latter, no depth perception emerged. In order to localize responsive areas, we ran an additional experiment during which a flickering checkerboard was presented to both eyes, interleaved with blank periods.

It is known that gradient-echo (GE) based sequences are considerably biased by macrovasculature, even at ultra-high field strengths[284]. Because of the high density of such large veins near the pial surface, layer-dependent BOLD signals acquired with GE sequences typically suffer of large biases towards superficial layers[264,285]. Spin-echo based sequences, such as 3D GRASE, have proved more immune to macrovasculature contributions and thus provide increase cortical specificity[284]. We chose the 3D GRASE sequence for this reason.

However, using 3D GRASE also imposes considerable restrictions on data acqui-

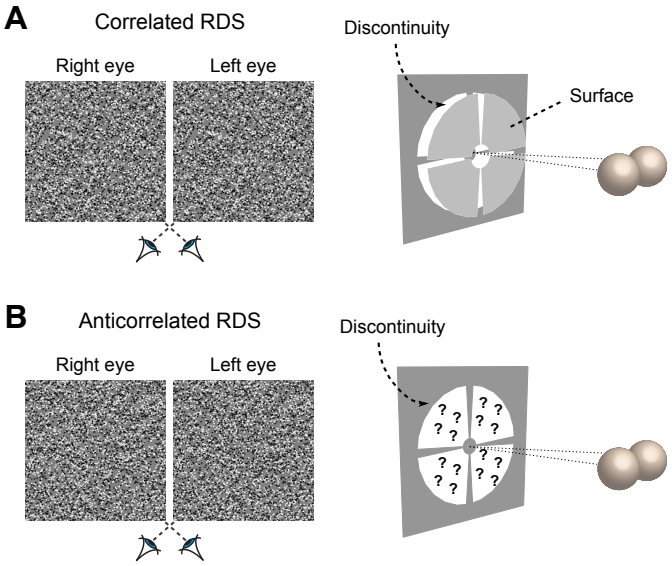


Figure 5.1: Correlated and anticorrelated stimuli used in the main experiment. (A) In correlated RDS, stimuli are rendered such that corresponding points across the left and right images have matching contrast polarity (i.e. a bright element in one eye is paired with a bright element in the other eye). In these displays, humans with normal binocular vision perceive a solid, 3-dimensional structure composed of four concentric wedges. (B) In anticorrelated RDS, corresponding points are rendered with opposite contrast (i.e. a bright element in one eye is paired with a dark element in the other eye). In contrast with correlated stimuli, anticorrelated RDS did not elicit depth perception.

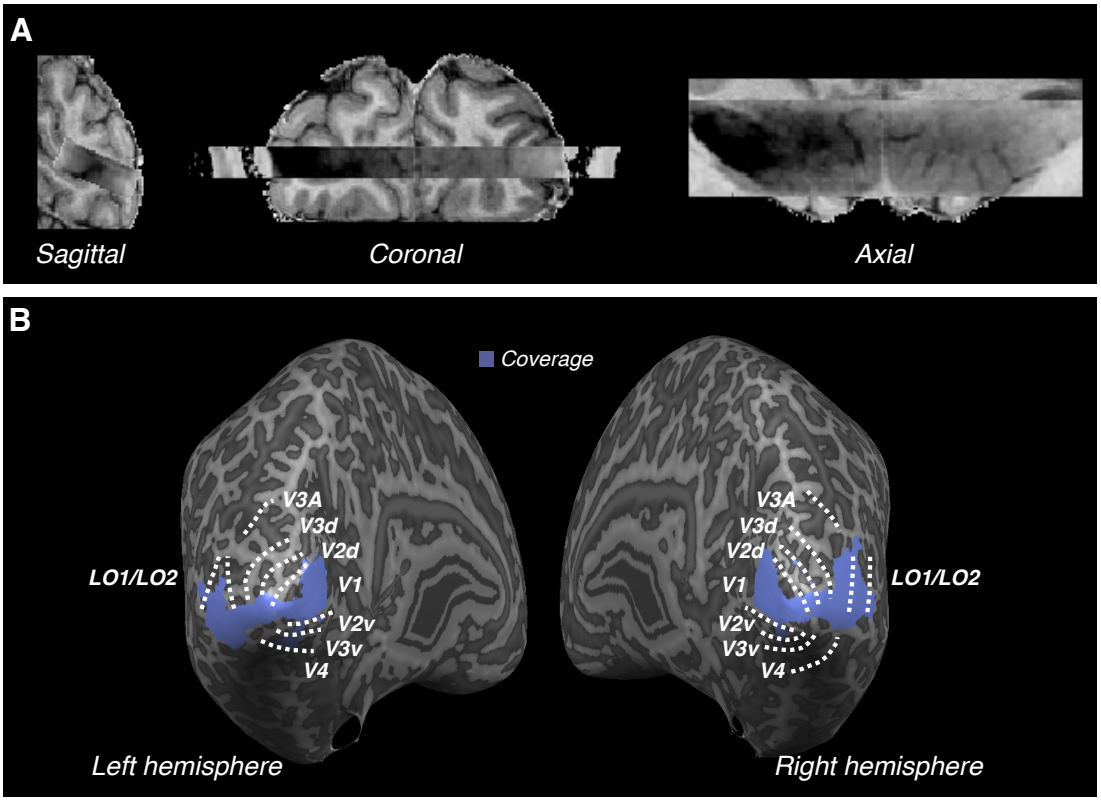


Figure 5.2: Imaging field-of-view for the main experiment. (A) We placed the acquisition slab parallel to the calcarine sulcus and attempted to image both hemispheres equally. (B) Previously collected retinotopic maps showed that we covered mainly V1 and lateral occipital areas LO1 and LO2.

sition and analysis. First and foremost, 3D GRASE offers less sensitivity to BOLD changes when compared to GE sequences[284]. Another worrying restriction is very limited coverage — while this is usually not an issue for conventional fMRI, the increase in resolution effectively decreases the volume of brain that can be covered by a particular number of slices. With 12 slices and a slice thickness of 0.8 mm, we were limited to covering 9.6 millimeters along the superior-inferior direction. We were mainly interested in imaging early visual areas, but the extended coverage in the left-right direction effectively allowed us to cover lateral occipital areas as well (Fig. 5.1). Slice positioning was guided by visual identification of the calcarine sulcus. Whenever possible, the slice was angled so as to cover both banks of the calcarine sulcus while maximizing coverage of lateral extrastriate areas in both hemispheres (Fig. 5.1).

Reduced coverage and high-resolution acquisitions also challenge off-the-shelf preprocessing tools. In particular, traditional algorithms for motion correction often fail

to converge to the correct solution. We have experienced this issue for all our participants, even though they typically moved less than 1 millimeter within an entire session. To compensate for this issue, we used a coarse-to-fine, exhaustive grid-search procedure for motion correction. We started by computing the similarity (here Pearson's correlation) between the reference and source scan for all possible points on a 6-dimensional grid (3 for translation and 3 for rotation) with five elements per dimension, and with each dimension spanning between ± 1 millimeter or degree. Thereafter, we chose the combination of parameters that produced the highest similarity, and then performed another iteration of grid-search over a smaller grid. This approach, although laborious, outperformed traditional algorithms for motion correction.

Despite the challenges, we obtained strong responses to stimulus delivery across striate and extrastriate areas (Fig. 5.2). Contrary to conventional fMRI data, our high-resolution protocol at ultra-high field yielded activity clusters that were well confined to the gray matter. Based on the F-statistic of the GLMs fit to the main experiment and to the localiser experiment, we identified two regions that were consistently driven by stimulus delivery across participants: a region of interest in the calcarine sulcus and another one in lateral occipital cortex. Retinotopy data confirmed that these two ROIs corresponded to areas V1 and LO (areas LO1/LO2), respectively.

Having confirmed the suitability of our imaging protocol, we then sought to examine the distribution of BOLD signal change across different cortical layers. To do so, we generated cortical grid meshes at different relative cortical depths between the white-matter/gray-matter boundary and the pial (Fig. 5.4A). This allowed us to assign a cortical depth label to each voxel within a region of interest. We did this via nearest-neighbour interpolation—that is, a particular voxel was assigned the relative cortical depth label corresponding the that of the closest cortical grid mesh. We then divided the cortical depth labels in three bins — deep, middle and superficial layers — and finally looked at the difference between activity evoked by correlated and anticorrelated RDS. In early visual cortex, we found little activity and no evident differences across layers (Fig. 5.4B). In contrast, we found a much greater effect in extrastriate areas, where a bias towards superficial layers was also evident. Thus, on the basis of univariate signals, we found a preference for stereopsis in lateral occipital cortex, but not in early visual cortex.

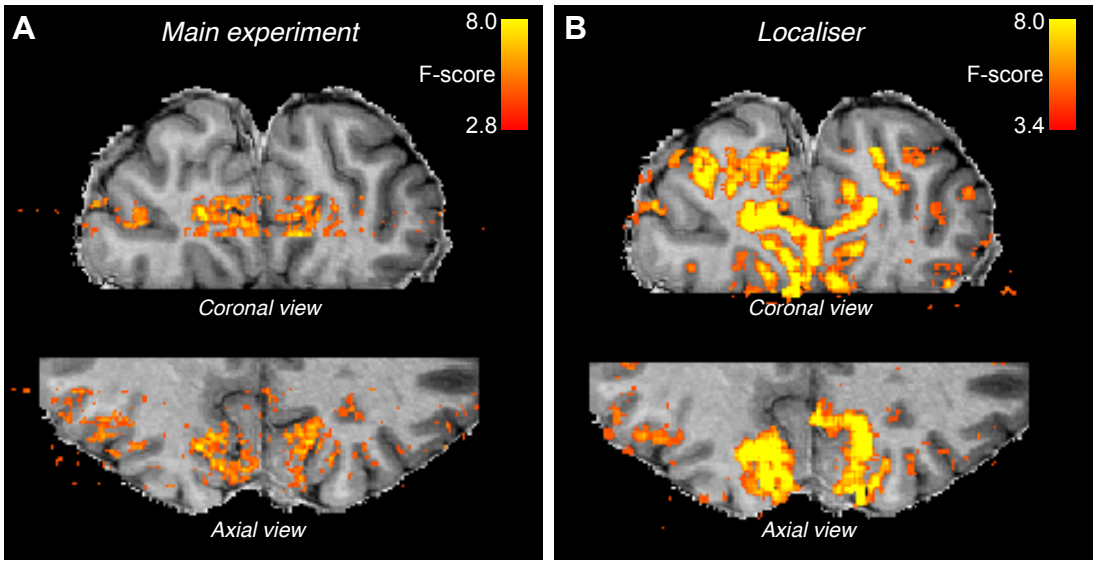


Figure 5.3: General linear modeling of task related activity demonstrated that our imaging protocol captured changes in signal associated with stimulus presentation, both in the main experiment (A) and in the localiser experiment (B).

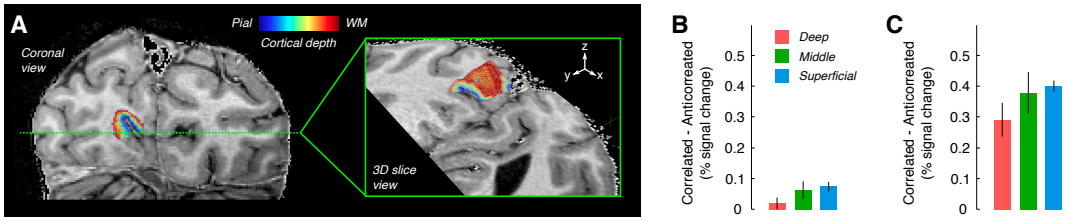


Figure 5.4: Layer-dependent activity in striate and extrastriate cortex (A). We used the Laplace layer sampling method[283] to obtain a cortical depth label for each voxel within the V1 and LO regions of interest. We found little difference in univariate V1 signal for correlated and anticorrelated stimuli (B). Conversely, we observe greater overall activation in LO relative to V1, and an apparent increase in layer-dependent activity towards superficial layers. Error bars denote standard error across participants.

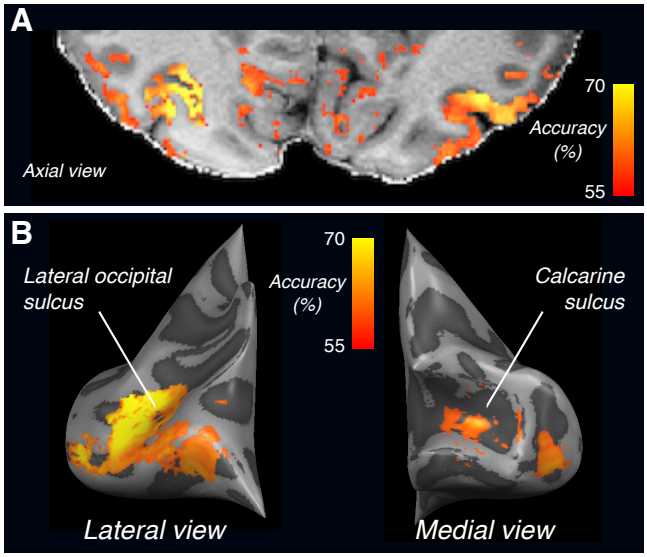


Figure 5.5: Searchlight classification for cRDS *vs* aRDS for a representative subject in volume (A) and cortical surface (B) space. We found above chance accuracy in primary visual cortex as well as lateral occipital areas.

Based on our previous analysis, concluding that activity in primary visual cortex was not related to stereoscopic perception would be premature. In particular, it is possible that fMRI measurements in V1 encapsulate stimulus related information across the activity of multiple instead of single voxels. To examine this possibility, we used support vector machines to decode binocular stimulus correlation (cRDS *vs.* aRDS) based on the pattern of activity across a population of voxels. Using a searchlight approach, we found that it was indeed possible to decode the stimulus class above chance level in both primary visual and lateral occipital cortex (Fig. 5.5).

Having established that both V1 and lateral occipital contain multivariate signals related to our experimental manipulation, we then asked whether such signals depend on cortical depth. To do so, we used the cortical depth sampling labels (see above) to subdivide voxels in three sub-ROIs: deep, middle and superficial layers. Consistent with our previous univariate analysis, we found a higher decoding accuracy in extrastriate relative to primary visual cortex and an apparent bias towards superficial layers (Fig. 5.6A,B; compare with Fig. 5.4B,C).

One possible concern with our previous approach is that binning voxels into 'deep', 'middle' and 'superficial' layers is somewhat arbitrary, and often biases the number of features (i.e. voxels) available for multivariate analysis (e.g. in the fundus

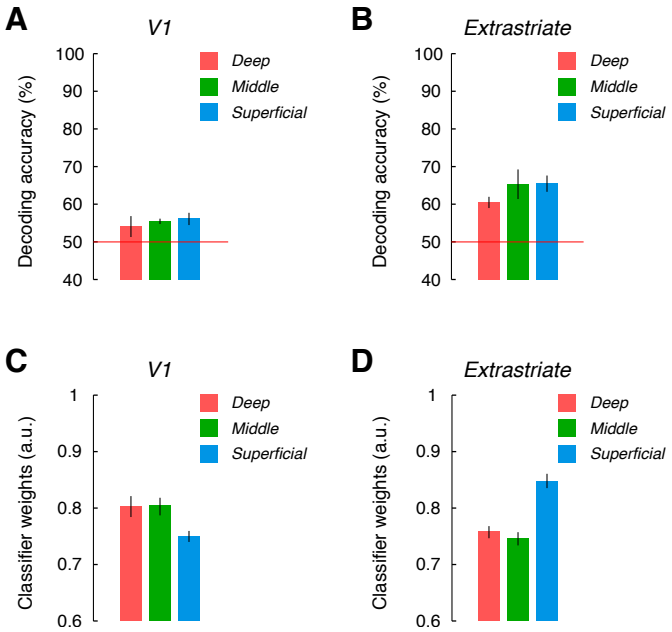


Figure 5.6: ROI based multivariate classification analysis. Classification accuracy (A, B) and classifier weights (C, D) per layer in V1 (A, C) and LO (B, D). These results indicate a bias towards superficial layers in LO, but not in V1. Error bars depict standard error across participants.

of a sulcus we may observe that more voxels are assigned to deep relative to superficial layers). Therefore, instead of subdividing each ROI in different layers, we trained classifiers using the entire pattern of voxels in an ROI, and then examined the weights attributed to voxels as a function of cortical depth. We found that the classifier assigned systematically higher weights to the superficial layers in lateral occipital areas (Fig. 5.6D), but that pattern was not observed for V1 (Fig. 5.6C).

5.4 Discussion

Here, we have examined how human visual cortex responds in the presence and absence of stable three-dimensional perception as measured using sub-millimeter fMRI. We manipulated perceptual relevance by using correlated and anticorrelated random-dot stereograms. Correlated stereograms could be easily fused to perceive a three-dimensional surface, while anticorrelated stimuli could not, and therefore did not elicit stable three-dimensional perception. Using ultra-high field imaging with isotropic sub-millimeter resolution, we investigated BOLD responses to these stimuli at differ-

ent cortical layers of primary and lateral-occipital visual cortex. We observed evident univariate signal change in LO during periods of stable three-dimensional perception, while less differential signals were found in primary visual cortex. Multivariate pattern classification analyses confirmed this dissociation between V1 and LO. Most importantly, our results suggest that ultra-high field fMRI is a valuable tool to examine layer-dependent activity associated with stereopsis.

Methodological limitations and neurovascular origin

Given that BOLD measurements are a haemodynamic proxy of neural activity, the right interpretation of our experimental results must also be a cautious one. We acquired BOLD signals at sub-millimeter resolution to investigate responses of neural populations in different cortical layers. Although the BOLD signal appears to be more locally registered to neural activity than previously thought[286], it is reasonable to assume that biases in the type and amount of vasculature across the cortical layers[287] may influence our results.

We have incorporated several measures to avoid vascular biases from interfering with the results. In particular, by scanning at ultra-high field, we greatly reduce macrovascular contributions[260–262]. Additionally, we chose to use a spin-echo based sequence (3D GRASE[288]) to provide increased spatial specificity — this sequence has an estimated point-spread function of approximately 0.5 mm (Gaussian FWHM) and is more immune to macrovascular contributions[284]. These characteristics of our imaging protocol are likely to have largely attenuated, if not eliminated, the large macrovascular biases that typically corrupt the BOLD signal.

To our knowledge, our study was the first attempt to examine how different cortical layers are involved in human stereoscopic perception. As we highlighted earlier, such endeavour posed significant challenges at the level of data acquisition and analysis. On the analysis side, reduced coverage and high-resolution data required the development of new preprocessing and analysis pipelines for fMRI data — some of which will keep improving as the community develops new tools and converges to best practices. For instance, methods for more accurate cortical depth sampling are being developed and are likely to become standard in the near future[289].

On the data acquisition end, we faced stringent limits on coverage. We followed

a quality by design approach by choosing 3D GRASE over a gradient-echo based sequence, since this provided us with the least opportunity for observing macrovascular driven effects[284]. Unfortunately, this entailed working with reduced coverage and lower signal-to-noise ratio. Future work might choose follow a quality by control approach instead. Using a gradient-echo based sequence will provide room for increasing coverage and will also yield higher signal-to-noise ratio: the former would allow us to simultaneously image many of the areas involved in stereopsis (potentially even the whole brain via multiband imaging); the latter would increase the likelihood of detecting small effects or potentially reduce the duration of each experiment, which effectively attenuates the impact of head motion. The main drawback would then relate to the interpretability of cortical profiles — the true pattern of layer-dependent activation might be masked by a mixture of signals poorly related to local neural activity. Finally, cerebral blood volume (CBV) measurements are also a promising technique that could be used for probing layer-dependent activity in the cortex. It has been shown that CBV measurements provide increase specificity and are therefore more interpretable than gradient-echo BOLD measurements[290]. However, as with 3D GRASE, CBV measurements impose tighter constraints on coverage.

An important pathway to establish the best protocol for imaging would be to validate layer-dependent fMRI measurements against observations made by neuroanatomists and neurophysiologists. A simple benchmark exists in the case of stereopsis: we know from a vast body of physiological work that neurons in layer 4C in V1 are monocular, and that binocularity increases towards supra- and infragranular layers. Therefore, we would expect little or no binocular information around intermediate cortical layers, and substantial binocular information in superficial and deep layers. However, that was not what we observed. Although we prioritized spatial specificity in choosing 3D GRASE, it is possible that the degree of specificity of the sequence is not high enough to detect such a U-shaped cortical profile — particularly if voxels in intermediate layers are contaminated by activity from supragranular and infragranular layers.

The role of cortical layers in stereoscopic vision

Our results suggest that activity in superficial layers of lateral occipital regions is a better correlate of stable stereoscopic perception than that of early visual areas (Figs.

5.4 and 5.6). While this is a novel and interesting finding, an explanatory account of the involvement of superficial layers of LO in stereoscopic processing cannot be built based on our data alone (or based on any fMRI dataset in general). However, we do find evidence for different profiles of activation across layers in primary visual and lateral occipital cortex. In what follows, we venture some potential explanations for our observations.

One possibility is that individual neurons in superficial layers of lateral occipital cortex are not selective for disparity in anticorrelated RDS. Neurophysiological investigations suggest that the response to anticorrelated stimuli decreases as we go up the cortical hierarchy[87,121], and eventually disappears in areas of the visual cortex specialized for object recognition[107]. Therefore, the preference for stereopsis that we observe in superficial layers of lateral occipital cortex could be a reflection of this feedforward process, which typically propagates through superficial layers of cortex.

Another possibility is that the nature of coding of stereoscopic information differs between primary visual and lateral occipital cortex. It is possible that stereoscopic information in primary visual cortex is dispersed across layers and is multivariate in nature. There is at least evidence against a univariate code because many individual V1 neurons have opposite disparity tuning curves for correlated and anticorrelated RDS[87]. On the other hand, neural activity at the level of individual neurons in LO might be more informative about stereo correspondence. Thus, it is possible that the differences in the laminar profile between V1 and LO reflect distinct coding principles rather than absence or presence of signals that support stereopsis.

Finally, it is also possible that the differences between the laminar profiles observed in V1 and LO are a consequence of the underlying cortical organization. In V1, it is known that neurons with different selectivity for binocular disparity are typically intermixed[144], and this lack of clustering could limit the detection of disparity selective responses as measured with fMRI. Conversely, as shown in the previous chapter, clustering for binocular disparity has been observed across many extrastriate areas, including area V3B/KO — an area located just next to LO. Thus, it is possible that an increase in clustering of disparity selective neurons in LO facilitated the detection of layer-dependent changes in this area, while the lack of clustering hindered the detection of a similar effect in V1.

Chapter 6

Discussion

In this thesis, I present theoretical and experimental studies that aim to advance our knowledge about the neural computations that support stereopsis. In what follows, I shall discuss the implications — and limitations — of our findings for our understanding of stereopsis.

Marr revisited, not revoked

One approach to investigate neural computations is to start with a formal examination of the computational goals for a particular task. This approach was first formulated by David Marr[1,158], and was at the heart of early theoretical work on stereopsis[151,158]. In the case of stereopsis, Marr divided the computational problem in two main parts: (i) establishing correspondence between the left and right images for each image element, and (ii) computing the binocular disparity between corresponding points. David Marr's formulation of the stereo computational problem was thus very intuitive. Having defined the computational problem (and some constraints), an algorithm design phase would then follow.

This approach is elegant and praised by many researchers to this date. However, it is not without its considerable pitfalls. In particular, an incorrect or incomplete specification of the computational problem will likely affect the subsequent level of analysis. A similar argument extends to the algorithmic level as well. As Minsky highlighted, one problem with the 'Marrian' approach is that it requires heavy feature hand-engineering, and very often the features that we come up with provide highly

suboptimal representations[291].

Our work too stems from thinking about computation in the first place, but we rely on neural networks to learn features that allow us to build better representations. In turn, this approach does require the loose definition of the building blocks of the algorithm that performs the computation. In our case, previous knowledge of the basic computational properties of disparity selective cells in V1 was instrumental in defining the building blocks of the neural network. In other cases (e.g. object recognition), defining such building blocks might be considerably more ambiguous because less is known about the properties and hierarchy of neurons involved in the computation of interest. Note, however, that the *a priori* definition of these building blocks does not implicate that the computation is performed by the precise architecture of the neural network. I argue that this approach — based on optimizing neural networks for particular neural computations — forms the basis for a new 'Marrian' approach for the machine learning era.

On binocular disparity

Strictly, the definition of binocular disparity — the difference between the positions of corresponding features in the left and right eyes — requires correspondence between the elements of the left and right images. Horace Barlow and colleagues[45] incorporated this idea in the interpretation of their findings that individual neurons in cat V1 have similar receptive fields in slightly different positions in the left and right eye: the similarity between the left and right receptive fields implied that they could be detecting the presence of similar features, while the difference in the RF position in the left and right eyes could encode binocular disparity. Later, this intuitive interpretation was found to be over-simplistic because many neurons have highly dissimilar receptive fields in the left and right eyes — instead of being disparate in their position, they are also very different, often antagonistic, in their phase[67]. This has long been regarded as a puzzle in the field.

Here I report that neurons optimized for estimating depth from disparity develop large phase disparities (i.e. tuned-inhibitory neurons). It is hard to see how neurons with such large phase disparities could look for similar elements across the eyes. In other words, how can such neurons explicitly solve the correspondence problem?

The responses of these neurons, which turn out to be the most informative to infer depth, do not seem to relate in any way to matching of similar features across the eyes. In this sense, I argue that they should not be thought of as neurons that attempt to explicitly solve the correspondence problem.

An apparent contradiction emerges at this point: how can a neuron not be related to solving the correspondence problem, but yet be very informative about the depth contained in disparate binocular images? The definition of binocular disparity requires correspondence. I argue that if we accept that tuned-inhibitory neurons do not play a role in determining stereo-correspondence, we should be prepared to accept that these neurons do not encode binocular disparity according to its strict definition — the positional difference between corresponding features in the left and right eyes. Because these neurons are the most informative for estimating depth, we should also be prepared to accept that binocular disparity — the *positional* difference between *corresponding* elements — might not be that important for depth perception. From this standpoint, these neurons seem to exploit *differences* (not only positional) between the left and right images as their cue to depth. This formulation allows us to bring together stereopsis with and without binocular correspondence.

What is the role of suppression?

Our theoretical and psychophysical work suggests that suppression may play an important role in stereopsis. Neurophysiologists have only recently started characterizing suppression in disparity selective cells in V1[80,81], but the data so far seem to support this conclusion[81]. Further work will be necessary to better understand the suppressive mechanisms involved, but the existent data and the results that we report here invite some speculation. Tanabe and Cumming found that suppression is only slightly delayed with respect to excitation, which would point to a fast, feedforward suppression mechanism — perhaps akin to that of cross-orientation suppression in V1[292]. This is in agreement with the predictions stemming from our theoretical work. The existence of a fast suppressive mechanism is also consistent with our psychophysical data, where we found a detrimental effect of introducing a very short onset asynchrony between signals that are thought to drive excitation and suppression of specific disparity detectors. However, our psychophysical results also indi-

cate the existence of a suppressive effect that is spatially more broad than the smallest excitatory effects. This could point to a different suppressive mechanism, perhaps mediated by slower and less precise feedback connections. Understanding the precise suppressive mechanisms requires further neurophysiological research, but on the basis of our data we speculate that two suppressive mechanisms at the level of V1 — a fast mechanism based on feedforward connections, and a slower one based on lateral or feedback connections.

Specialization for stereopsis

Although nearly every experimental technique has been used to investigate stereopsis, the field has not been able to converge on a single specialized area for stereopsis. Instead, the evidence so far points to distributed coding across many cortical areas, mainly in the ventral and dorsal visual streams. Here I report evidence of systematic cortical organization for depth from binocular disparity in areas V3A and V3B/KO. However, it is possible that similar organization is present elsewhere in visual cortex — perhaps in the ventral stream, which we were unable to image due to field-of-view limitations. Furthermore, we have yet to characterize this cortical organization: we show that the disparity preferences are persistently represented in the cortex, but we were not able to identify the rules that govern the spatial arrangement of disparity preferences. It would be interesting to compare the results obtained in the dorsal stream with preference maps for the ventral stream, which would hopefully help to further dissociate the role of the ventral and dorsal streams in stereoscopic processing.

Another question that remains to be answered is concerned with the purpose of such cortical organization. Previous work suggests that cortical organization might be intimately related to neural activity that correlates with perception on a trial-by-trial basis[91]. We were unable to test this with fMRI due to the limited temporal resolution. Based on the literature, it seems that a correlation exists between the presence of cortical organization and neural signals related to depth perception[90,91,99,102,117,118,293].

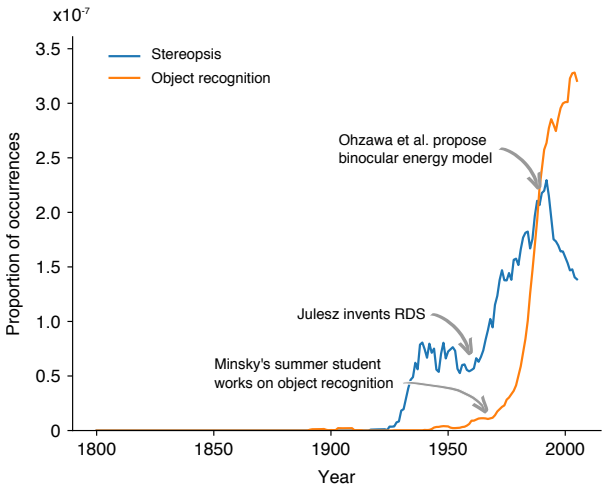


Figure 6.1: Popularity of the n-grams 'stereopsis' and 'object recognition'.

Future work

Until the end of the 1990's, *Nature* and *Science* were often filled with reports of behavioural and neurophysiological studies concerned with stereopsis; in contrast, little research on stereopsis has caught the eye of such high impact journals in the last 15 years or so. Let us consider the frequency of the n-grams 'stereopsis' and 'object recognition' in the database of the Ngram Viewer project (Fig. 6.1), which contains over 5 million books (approximately 4% of all the books ever published)[294]. The term 'stereopsis' increases in frequency first around the 1920's (likely due to the popularity of plasticon cinemas), and then again following the invention of the random-dot stereogram. Consistent with the trend observed in high-impact publishing, a worrying decrease in the frequency of the n-gram 'stereopsis' has been observed since the 1990's. Conversely, the n-gram 'object recognition' has been steadily increasing in popularity approximately since Marvin Minsky hired a student to work on a summer project with the goal of solving object recognition. Only after this point was object recognition considered an interesting problem.

To a non expert, figure 6.1 might suggest that stereopsis is already well understood or that it is no longer an interesting research topic. In what follows I will argue that this is not the case. We still have a poor understanding of how different brain regions are involved stereoscopic vision. Beyond primary visual cortex, the scientific

community has not yet managed to converge to a general computational framework for stereopsis, let alone designing mechanistic models that are able to predict neural responses to naturalistic 3D stimuli. Characterizing the contributions of different cortical layers is one promising line of research for exploring the interactions between different visual areas involved in stereopsis. In particular, it would be interesting to exploit the exploratory power of high-resolution functional magnetic resonance imaging to identify key circuits, which could then be dissected in greater detail using layer-specific electrical or optical recordings. The interactions between vergence, accommodation and stereopsis at the neural level are also poorly understood. There is evidence that disparity selectivity in V1 is greatly modulated by viewing distance[295], but, to our knowledge, models of disparity selectivity in V1 have yet to explain these findings.

Additionally, down-weighting the importance that studying stereopsis can have in advancing our understanding of the brain is a mistake. First, as I have mentioned before, stereopsis seems to rely on multiple brain regions across the entire brain, which attests its suitability to study how different brain regions work together to support perception. Second, stereopsis is a classical demonstration of how the brain excels at rapidly doing inverse graphics. Understanding how the brain achieves this may provide valuable insights to the fields of machine learning and artificial intelligence — in the same way the connectionist principles inspired the development of modern, state-of-the-art artificial intelligence.

References

[1] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc., New York, NY, USA, 1982. 1, 139

[2] I. Howard and B. Rogers. *Seeing in Depth*. Oxford University Press, 2008. 1, 4

[3] J. M. Allman. *Evolving brains*, volume no. 68. Scientifc American Library, New York, 1999. 2

[4] M. Cartmill. New views on primate origins. *Evolutionary Anthropology: Issues, News, and Reviews*, 1(3):105–111, 1992. 2

[5] R. D. Martin and A.-E. Martin. *Primate origins and evolution: a phylogenetic reconstruction*. Princeton University Press, Princeton, N.J., 1990. 2

[6] W. E. L. G. Clark. *Early forerunners of man: a morphological study of the evolutionary origin of the primates*. W. Wood and Company, Baltimore, 1934. 2

[7] B. Julesz. *Foundations of cyclopean perception*. University of Chicago Press, Chicago, 1971. 2, 16, 25

[8] M. A. Changizi and S. Shimojo. "x-ray vision" and the evolution of forward-facing eyes. *J Theor Biol*, 254(4):756–67, Oct 2008. 2

[9] S. Jainta, H. I. Blythe, and S. P. Liversedge. Binocular advantages in reading. *Curr Biol*, 24(5):526–30, Mar 2014. 2

[10] D. R. Melmoth and S. Grant. Advantages of binocular vision for the control of reaching and grasping. *Experimental Brain Research*, 171(3):371–388, 2006. 2

[11] K. Nakayama and S. Shimojo. da vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Res*, 30(11):1811–25, 1990. 5, 7, 8, 15, 18, 34

[12] C. Wheatstone. Contributions to the physiology of vision.–part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128:371–394, 01 1838. 6, 15, 25, 31, 67, 124

[13] H. W. Dove. Bericht über die zur bekanntmachung geeigneten verhandlungen der königlich preußischen akademie der wissenschaften zu berlin. ueber die combination der eindruecke beider ohren und beider augen zu einem eindrueck. *Berlin: Verl. d. Kgl. Akad. d. Wiss.*, pages 251–2, 1841. 6

[14] H. W. Dove. Ueber stereoskopie. *Annalen der Physik*, 186(7):494–498, 1860. 6

[15] F. Donders. Das binoculare sehen und die vorstellung von der dritten dimension. *Archiv für Ophthalmologie*, 13(1):1–48, 1867. 6

[16] B. Julesz. Binocular depth perception without familiarity cues. *Science*, 145(3630):356–62, Jul 1964. 6, 7, 8, 67, 83

[17] L. Kaufman. On the nature of binocular disparity. *Am J Psychol*, 77:393–402, Sep 1964. 6, 15, 16

[18] L. Kaufman and C. Pitblado. Further observations on the nature of effective binocular disparities. *Am J Psychol*, 78:379–91, Sep 1965. 15, 16

[19] J. E. Mayhew and J. P. Frisby. Rivalrous texture stereograms. *Nature*, 264(5581):53–6, Nov 1976. 6, 15, 18

[20] M. Kaye. Stereopsis without binocular correlation. *Vision Res*, 18(8):1013–22, 1978. 6, 8, 15, 18

[21] L. M. Wilcox, J. M. Harris, and S. P. McKee. The role of binocular stereopsis in monoptic depth perception. *Vision Res*, 47(18):2367–77, Aug 2007. 6, 8

[22] B. Gillam and E. Borsting. The role of monocular regions in stereoscopic displays. *Perception*, 17(5):603–8, 1988. 7, 18, 34

[23] S. Shimojo, G. H. Silverman, and K. Nakayama. An occlusion-related mechanism of depth perception based on motion and interocular sequence. *Nature*, 333(6170):265–8, May 1988.

[24] S. Shimojo and K. Nakayama. Real world occlusion constraints and binocular rivalry. *Vision Res*, 30(1):69–80, 1990. 7, 8, 18

[25] B. Gillam, M. Cook, and S. Blackburn. Monocular discs in the occlusion zones of binocular surfaces do not have quantitative depth–a comparison with panum's limiting case. *Perception*, 32(8):1009–19, 2003. 7

[26] I. Tsirlin, L. M. Wilcox, and R. S. Allison. da vinci decoded: does da vinci stereopsis rely on disparity? *J Vis*, 12(12), 2012. 7

[27] M. J. Pianta and B. J. Gillam. Paired and unpaired features can be equally effective in human depth perception. *Vision Res*, 43(1):1–6, Jan 2003. 7

[28] I. Tsirlin, L. M. Wilcox, and R. S. Allison. Disparity biasing in depth from monocular occlusions. *Vision Res*, 51(14):1699–711, Jul 2011. 8

[29] J. M. Harris and L. M. Wilcox. The role of monocularly visible regions in depth and surface perception. *Vision Res*, 49(22):2666–85, Nov 2009. 8

[30] R. B. Lawson and W. L. Gulick. Stereopsis and anomalous contour. *Vision Res*, 7(3):271–97, Mar 1967. 8

[31] B. L. Anderson and K. Nakayama. Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychol Rev*, 101(3):414–45, Jul 1994. 16, 18, 34

[32] B. L. Anderson. The role of partial occlusion in stereopsis. *Nature*, 367(6461):365–8, Jan 1994. 8, 18

[33] B. G. Cumming, S. E. Shapiro, and A. J. Parker. Disparity detection in anti-correlated stereograms. *Perception*, 27(11):1367–77, 1998. 8, 13, 68, 83, 84

[34] A. I. Cogan, A. J. Lomakin, and A. F. Rossi. Depth in anticorrelated stereograms: effects of spatial density and interocular delay. *Vision Res*, 33(14):1959–75, Sep 1993. 8, 68, 83

[35] S. Tanabe, S. Yasuoka, and I. Fujita. Disparity-energy signals in perceived stereoscopic depth. *J Vis*, 8(3):22.1–10, 2008. 8

[36] T. Doi, S. Tanabe, and I. Fujita. Matching and correlation computations in stereoscopic depth perception. *J Vis*, 11(3):1, 2011. 8, 41, 83

[37] P. B. Hibbard, K. C. Scott-Brown, E. C. Haigh, and M. Adrain. Depth perception not found in human observers for static or dynamic anti-correlated random dot stereograms. *PloS one*, 9(1):e84087, Jan 2014. 8, 13, 68, 83

[38] J. C. Read and R. A. Eagle. Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision Res*, 40(24):3345–58, 2000. 8, 68, 83

[39] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts. What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 47(11):1940–1951, Nov 1959. 9, 38

[40] W. R. Levick. Receptive fields and trigger features of ganglion cells in the visual streak of the rabbits retina. *J Physiol*, 188(3):285–307, Feb 1967.

[41] H. B. Barlow. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1(4):371–94, 1972. 9

[42] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *J Physiol*, 148:574–91, Oct 1959. 9, 10, 38

[43] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160:106–54, Jan 1962. 10, 11, 14, 20

[44] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *J Physiol*, 195(1):215–43, Mar 1968. 9, 10, 11, 12, 14

[45] H. B. Barlow, C. Blakemore, and J. D. Pettigrew. The neural mechanism of binocular depth discrimination. *J Physiol*, 193(2):327–42, Nov 1967. 9, 11, 124, 140

[46] T. Nikara, P. O. Bishop, and J. D. Pettigrew. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp Brain Res*, 6(4):353–72, 1968.

[47] J. D. Pettigrew, T. Nikara, and P. O. Bishop. Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Exp Brain Res*, 6(4):391–410, 1968. 9, 11, 124

[48] J. D. Pettigrew and M. Konishi. Neurons selective for orientation and binocular disparity in the visual wulst of the barn owl (tyto alba). *Science*, 193(4254):675–8, Aug 1976. 9

[49] B. Scholl, J. Burge, and N. J. Priebe. Binocular integration and disparity selectivity in mouse primary visual cortex. *J Neurophysiol*, 109(12):3013–24, Jun 2013. 9

[50] D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex*, 1(1):1–47, January 1991. 9, 10, 121

[51] D. Y. Tsao, W. Vandluffel, Y. Sasaki, D. Fize, T. A. Knutsen, J. B. Mandeville, L. L. Wald, A. M. Dale, B. R. Rosen, D. C. Van Essen, M. S. Livingstone, G. A. Orban, and R. B. H. Tootell. Stereopsis activates v3a and caudal intraparietal areas in macaques and humans. *Neuron*, 39(3):555–568, July 2003. 10, 14, 24, 87, 100, 124

[52] J. J. Nassi and E. M. Callaway. Parallel processing strategies of the primate visual system. *Nature Reviews Neuroscience*, 10(5):360–372, April 2009. 9

[53] E. M. Callaway. Local circuits in primary visual cortex of the macaque monkey. *Annual review of neuroscience*, 21:47–74, 1998. 10

[54] K. D. Harris and T. D. Mrsic-Flogel. Cortical connectivity and sensory coding. *Nature*, 503(7474):51–8, Nov 2013. 9

[55] G. F. Poggio. Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative ophthalmology*, 11(5):368–377, May 1972. 10

[56] A. Burkhalter and D. C. Van Essen. Processing of color, form and disparity information in visual areas vp and v2 of ventral extrastriate cortex in the macaque monkey. *J Neurosci*, 6(8):2327–51, Aug 1986. 10

[57] J. H. R. Maunsell and W. T. Newsome. Visual processing in monkey extrastriate cortex. *Annual Review of Neuroscience*, 10(1):363–401, 2015/10/13 1987. 10

[58] G. F. Poggio and B. Fischer. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurophysiol*, 40(6):1392–405, Nov 1977. 11, 12, 13, 15, 95, 111, 124

[59] G. F. Poggio and W. H. Talbot. Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J Physiol*, 315:469–92, Jun 1981. 11

[60] A. Anzai, I. Ohzawa, and R. D. Freeman. Neural mechanisms for processing binocular information ii. complex cells. *Journal of neurophysiology*, 82(2):909–924, Aug 1999. 11, 12

[61] B. Cumming. Stereopsis: how the brain sees depth. *Curr Biol*, 7(10):R645–7, Oct 1997.

[62] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman. Encoding of binocular disparity by complex cells in the cat's visual cortex. *J Neurophysiol*, 77(6):2879–909, Jun 1997. 12

[63] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 249(4972):1037–41, Aug 1990. 11, 20, 26, 27, 28, 40, 55, 57, 68, 124

[64] I. Ohzawa and R. D. Freeman. The binocular organization of complex cells in the cat's visual cortex. *J Neurophysiol*, 56(1):243–59, Jul 1986. 11, 12

[65] R. von der Heydt, C. Adorjani, P. Hänny, and G. Baumgartner. Disparity sensitivity and receptive field incongruity of units in the cat striate cortex. *Exp Brain Res*, 31(4):523–45, Apr 1978. 11

[66] D. Ferster. A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *J Physiol*, 311:623–55, Feb 1981. 11, 54

[67] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, 352(6331):156–9, Jul 1991. 11, 26, 27, 29, 39, 68, 140

[68] A. Anzai, I. Ohzawa, and R. D. Freeman. Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proc Natl Acad Sci U S A*, 94(10):5438–43, May 1997. 11

[69] A. Anzai, I. Ohzawa, and R. D. Freeman. Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J Neurophysiol*, 82(2):874–90, Aug 1999. 11

[70] A. Anzai, I. Ohzawa, and R. D. Freeman. Neural mechanisms for processing binocular information i. simple cells. *J Neurophysiol*, 82(2):891–908, Aug 1999. 55, 56

[71] D. Y. Tsao, B. R. Conway, and M. S. Livingstone. Receptive fields of disparity-tuned simple cells in macaque v1. *Neuron*, 38(1):103–14, Apr 2003. 11, 26, 27, 54, 68

[72] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, 352(6331):156–159, Jul 1991.

[73] S. J. D. Prince, B. G. Cumming, and P. A. J. Range and mechanism of encoding of horizontal disparity in macaque v1. *Journal of Neurophysiology*, 87:209–221, 2002. 11, 12, 26, 27, 68, 95, 111, 123

[74] M. S. Livingstone and D. Y. Tsao. Receptive fields of disparity-selective neurons in macaque striate cortex. *Nat Neurosci*, 2(9):825–32, Sep 1999. 12, 23

[75] K. A. Archie and B. W. Mel. A model for intradendritic computation of binocular disparity. *Nat Neurosci*, 3(1):54–63, Jan 2000. 12, 23

[76] T. Duong, B. D. Moore, 4th, and R. D. Freeman. Adaptation changes stereoscopic depth selectivity in visual cortex. *J Neurosci*, 31(34):12198–207, Aug 2011. 12

[77] K. S. Sasaki, Y. Tabuchi, and I. Ohzawa. Complex cells in the cat striate cortex have multiple disparity detectors in the three-dimensional binocular receptive fields. *J. Neurosci.*, 30(41):13826–13837, October 2010. 12, 40, 85

[78] M. Baba, K. S. Sasaki, and I. Ohzawa. Integration of multiple spatial frequency channels in disparity-sensitive neurons in the primary visual cortex. *J Neurosci*, 35(27):10025–38, Jul 2015. 40, 60

[79] D. Kato, M. Baba, K. S. Sasaki, and I. Ohzawa. Effects of generalized pooling on binocular disparity selectivity of neurons in the early visual cortex. *Philos Trans R Soc Lond B Biol Sci*, 371(1697), Jun 2016. 12, 59, 60, 85

[80] S. Tanabe and B. G. Cumming. Delayed suppression shapes disparity selective responses in monkey v1. *J Neurophysiol*, 111(9):1759–69, May 2014. 12, 40, 80, 84, 85, 141

[81] S. Tanabe, R. M. Haefner, and B. G. Cumming. Suppressive mechanisms in monkey v1 help to solve the stereo correspondence problem. *J Neurosci*, 31(22):8295–305, Jun 2011. 12, 34, 40, 80, 85, 141

[82] R. M. Haefner and B. G. Cumming. Adaptation to natural binocular disparities in primate v1 explained by a generalized energy model. *Neuron*, 57(1):147–58, Jan 2008. 12, 23, 29, 40, 84

[83] G. F. Poggio, F. Gonzalez, and F. Krause. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J Neurosci*, 8(12):4531–50, Dec 1988. 12, 95, 111, 113, 123

[84] S. R. Lehky and T. J. Sejnowski. Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *J Neurosci*, 10(7):2281–99, Jul 1990. 12, 76, 83, 98, 99, 112, 117, 120, 123

[85] S. R. Lehky, A. Pouget, and T. J. Sejnowski. Neural models of binocular depth perception. *Cold Spring Harbor symposia on quantitative biology*, 55:765–777, 1990. 12

[86] G. F. Poggio and T. Poggio. The analysis of stereopsis. *Annu Rev Neurosci*, 7:379–412, 1984. 12

[87] B. G. Cumming and A. J. Parker. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, 389(6648):280–3, Sep 1997. 13, 22, 28, 29, 51, 63, 68, 83, 124, 138

[88] B. G. Cumming and A. J. Parker. Local disparity not perceived depth is signaled by binocular neurons in cortical area v1 of the macaque. *J Neurosci*, 20(12):4758–67, Jun 2000. 13

[89] J. M. Samonds, B. R. Potetz, C. W. Tyler, and T. S. Lee. Recurrent connectivity can account for the dynamics of disparity processing in v1. *J Neurosci*, 33(7):2934–46, Feb 2013. 13, 28, 29, 40, 51, 63, 83, 84

[90] H. Nienborg and B. G. Cumming. Macaque v2 neurons, but not v1 neurons, show choice-related activity. *J Neurosci*, 26(37):9567–78, Sep 2006. 13, 142

[91] H. Nienborg and B. G. Cumming. Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. *J Neurosci*, 34(10):3579–85, Mar 2014. 13, 15, 142

[92] G. S. Masson, C. Busettini, and F. A. Miles. Vergence eye movements in response to binocular disparity without depth perception. *Nature*, 389(6648):283–6, Sep 1997. 13, 15, 18

[93] F. Gonzalez and R. Perez. Neural mechanisms underlying stereoscopic vision. *Prog. Neurobiol.*, 55(3):191–224, June 1998. 13

[94] P. Neri. A stereoscopic look at visual cortex. *J. Neurophysiol.*, 93(4):1823–1826, April 2005.

[95] A. J. Parker. Binocular depth perception and the cerebral cortex. *Nat. Rev. Neurosci.*, 8(5):379–391, May 2007. 13, 14, 124

[96] D. H. Hubel and T. N. Wiesel. Stereoscopic vision in macaque monkey. cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature*, 225(5227):41–2, Jan 1970. 13, 15

[97] S. M. Zeki. The third visual complex of rhesus monkey prestriate cortex. *The Journal of physiology*, 277:245–272, April 1978. 13

[98] A. Anzai, S. A. Chowdhury, and G. C. DeAngelis. Coding of stereoscopic depth information in visual areas v3 and v3a. *J Neurosci*, 31(28):10270–82, Jul 2011. 13, 14, 15, 87, 95, 100, 119, 121

[99] G. C. DeAngelis, B. G. Cumming, and W. T. Newsome. Cortical area mt and the perception of stereoscopic depth. *Nature*, 394(6694):677–680, August 1998. 13, 142

[100] G. C. DeAngelis and W. T. Newsome. Organization of disparity-selective neurons in macaque area mt. *Journal of Neuroscience*, 19(4):1398–1415, 1999. 13, 15, 87, 104, 111, 119

[101] S. Eifuku and R. H. Wurtz. Response to motion in extrastriate area mstl: disparity sensitivity. *J Neurophysiol*, 82(5):2462–75, Nov 1999. 13

[102] H. M. Shiozaki, S. Tanabe, T. Doi, and I. Fujita. Neural activity in cortical area v4 underlies fine disparity discrimination. *J Neurosci*, 32(11):3830–41, Mar 2012. 13, 14, 124, 142

[103] K. Umeda, S. Tanabe, and I. Fujita. Representation of stereoscopic depth based on relative disparity in macaque area v4. *J Neurophysiol*, 98(1):241–52, Jul 2007. 14

[104] S. Tanabe, T. Doi, K. Umeda, and I. Fujita. Disparity-tuning characteristics of neuronal responses to dynamic random-dot stereograms in macaque visual area v4. *J Neurophysiol*, 94(4):2683–99, Oct 2005.

[105] M. Watanabe, H. Tanaka, T. Uka, and I. Fujita. Disparity-selective neurons in area v4 of macaque monkeys. *J Neurophysiol*, 87(4):1960–73, Apr 2002. 13, 14

[106] T. Uka, H. Tanaka, K. Yoshiyama, M. Kato, and I. Fujita. Disparity selectivity of neurons in monkey inferior temporal cortex. *J Neurophysiol*, 84(1):120–32, Jul 2000. 13

[107] P. Janssen, R. Vogels, Y. Liu, and G. A. Orban. At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron*, 37(4):693–701, Feb 2003. 13, 14, 30, 84, 138

[108] P. Janssen, R. Vogels, and G. A. Orban. Three-dimensional shape coding in inferior temporal cortex. *Neuron*, 27(2):385–97, Aug 2000. 14

[109] P. Janssen, R. Vogels, and G. A. Orban. Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proc Natl Acad Sci U S A*, 96(14):8217–22, Jul 1999. 13, 14

[110] B.-E. Verhoef, P. Michelet, R. Vogels, and P. Janssen. Choice-related activity in the anterior intraparietal area during 3-d structure categorization. *J Cogn Neurosci*, 27(6):1104–15, Jun 2015. 13

[111] B.-E. Verhoef, R. Vogels, and P. Janssen. Contribution of inferior temporal and posterior parietal activity to three-dimensional shape perception. *Curr Biol*, 20(10):909–13, May 2010.

[112] S. Srivastava, G. A. Orban, P. A. De Mazière, and P. Janssen. A distinct representation of three-dimensional shape in macaque anterior intraparietal area: fast, metric, and coarse. *J Neurosci*, 29(34):10613–26, Aug 2009. 13, 14

[113] J. W. Gnadt and L. E. Mays. Neurons in monkey parietal area lip are tuned for eye-movement parameters in three-dimensional space. *J Neurophysiol*, 73(1):280–97, Jan 1995. 13

[114] A. Genovesio and S. Ferraina. Integration of retinal disparity and fixation-distance related signals toward an egocentric coding of distance in the posterior parietal cortex of primates. *J Neurophysiol*, 91(6):2670–84, Jun 2004. 13

[115] C. L. Colby, J. R. Duhamel, and M. E. Goldberg. Ventral intraparietal area of the macaque: anatomic location and visual response properties. *J Neurophysiol*, 69(3):902–14, Mar 1993. 13

[116] S. Ferraina, M. Paré, and R. H. Wurtz. Disparity sensitivity of frontal eye field neurons. *J Neurophysiol*, 83(1):625–9, Jan 2000. 13

[117] S. Clery, B. Cumming, and H. Nienborg. Decision-related activity in v2 for a fine disparity discrimination task. *J Vis*, 15(12):830, 2015. 13, 124, 142

[118] H. Nienborg and B. G. Cumming. Psychophysically measured task strategy for disparity discrimination is reflected in v2 neurons. *Nat Neurosci*, 10(12):1608–14, Dec 2007. 13, 124, 142

[119] A. Smolyanskaya, R. M. Haefner, S. G. Lomber, and R. T. Born. A modality-specific feedforward component of choice-related activity in mt. *Neuron*, 87(1):208–19, Jul 2015. 13, 14

[120] B.-E. Verhoef, R. Vogels, and P. Janssen. Inferotemporal cortex subserves three-dimensional structure categorization. *Neuron*, 73(1):171–82, Jan 2012. 13, 14

[121] S. Tanabe, K. Umeda, and I. Fujita. Rejection of false matches for binocular correspondence in macaque visual cortical area v4. *J Neurosci*, 24(37):8170–80, Sep 2004. 13, 30, 51, 138

[122] K. Krug, B. G. Cumming, and A. J. Parker. Comparing perceptual signals of single v5/mt neurons in two binocular depth tasks. *J Neurophysiol*, 92(3):1586–96, Sep 2004. 13

[123] A. Takemura, Y. Inoue, K. Kawano, C. Quaia, and F. A. Miles. Single-unit activity in cortical area mst associated with disparity-vergence eye movements: evidence for population coding. *J Neurophysiol*, 85(5):2245–66, May 2001. 13

[124] O. M. Thomas, B. G. Cumming, and A. J. Parker. A specialization for relative disparity in v2. *Nat Neurosci*, 5(5):472–8, May 2002. 13, 14

[125] J. D. Nguyenkim and G. C. DeAngelis. Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J Neurosci*, 23(18):7117–28, Aug 2003. 14

[126] J. P. Roy, H. Komatsu, and R. H. Wurtz. Disparity sensitivity of neurons in monkey extrastriate area mst. *J Neurosci*, 12(7):2478–92, Jul 1992. 14

[127] J. P. Roy and R. H. Wurtz. The role of disparity-sensitive cortical neurons in signalling the direction of self-motion. *Nature*, 348(6297):160–2, Nov 1990. 14

[128] A. Rosenberg, N. J. Cowan, and D. E. Angelaki. The visual representation of 3d object orientation in parietal cortex. *J Neurosci*, 33(49):19352–61, Dec 2013. 14

[129] B. T. Backus, D. J. Fleet, A. J. Parker, and D. J. Heeger. Human cortical activity correlates with stereoscopic depth perception. *J Neurophysiol*, 86(4):2054–68, Oct 2001. 14, 24, 87, 100, 121, 124

[130] H. Ban and A. E. Welchman. fmri analysis-by-synthesis reveals a dorsal hierarchy that extracts surface slant. *J Neurosci*, 35(27):9823–35, Jul 2015. 14, 34

[131] C. W. Tyler. A stereoscopic view of visual processing streams. *Vision Res*, 30(11):1877–95, 1990. 14

[132] P. H. Schiller, N. K. Logothetis, and E. R. Charles. Functions of the colour-opponent and broad-band channels of the visual system. *Nature*, 343(6253):68–70, Jan 1990. 14

[133] T. Uka and G. C. DeAngelis. Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *J Neurosci*, 23(8):3515–30, Apr 2003. 14

[134] T. Uka. Linking neural representation to function in stereoscopic depth perception: Roles of the middle temporal area in coarse versus fine disparity discrimination. *Journal of Neuroscience*, 26(25):6791–6802, Jun 2006. 14

[135] A. W. Roe, A. J. Parker, R. T. Born, and G. C. DeAngelis. Disparity channels in early vision. *Journal of Neuroscience*, 27(44):11820–11831, Oct 2007. 14

[136] G. A. Orban, P. Janssen, and R. Vogels. Extracting 3d structure from disparity. *Trends Neurosci*, 29(8):466–73, Aug 2006. 14, 124

[137] C. R. Ponce, S. G. Lomber, and R. T. Born. Integrating motion and depth via parallel pathways. *Nature Neuroscience*, 11(2):216–223, February 2008. 14

[138] D. H. Hubel and T. N. Wiesel. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J Comp Neurol*, 158(3):267–93, Dec 1974. 14, 86

[139] T. N. Wiesel and D. H. Hubel. Ordered arrangement of orientation columns in monkeys lacking visual experience. *J Comp Neurol*, 158(3):307–18, Dec 1974.

[140] D. H. Hubel, T. N. Wiesel, and M. P. Stryker. Anatomical demonstration of orientation columns in macaque monkey. *J Comp Neurol*, 177(3):361–80, Feb 1978.

[141] D. H. Hubel and T. N. Wiesel. Anatomical demonstration of columns in the monkey striate cortex. *Nature*, 221(5182):747–50, Feb 1969. 14

[142] J. C. Horton and D. L. Adams. The cortical column: a structure without a function. *Philos. Trans. R. Soc. B-Biol. Sci.*, 360(1456):837–862, April 2005. 15

[143] S. LeVay and T. Voigt. Ocular dominance and disparity coding in cat visual cortex. *Vis Neurosci*, 1(4):395–414, 1988. 15, 87, 121

[144] S. J. D. Prince, A. D. Pointon, B. G. Cumming, and A. J. Parker. Quantitative analysis of the responses of v1 neurons to horizontal disparity in dynamic random-dot stereograms. *J Neurophysiol*, 87(1):191–208, Jan 2002. 15, 87, 121, 138

[145] D. H. Hubel and M. S. Livingstone. Segregation of form, color, and stereopsis in primate area 18. *J Neurosci*, 7(11):3378–415, Nov 1987. 15, 87

[146] G. Chen, H. D. Lu, and A. W. Roe. A map for horizontal disparity in monkey v2. *Neuron*, 58(3):442–50, May 2008. 15, 121

[147] P. Kara and J. D. Boyd. A micro-architecture for binocular disparity and ocular dominance in visual cortex. *Nature*, 458(7238):627–31, Apr 2009. 15, 87

[148] D. Adams and S. Zeki. Functional organization of macaque V3 for stereoscopic depth. *Journal of neurophysiology*, 86(5):2195–2203, 2001. 15, 87, 121

[149] D. H. Hubel, T. N. Wiesel, E. M. Yeagle, R. Lafer-Sousa, and B. R. Conway. Binocular stereoscopy in visual areas v-2, v-3, and v-3a of the macaque monkey. *Cerebral Cortex*, Oct 2013. 15, 87, 100, 119

[150] B. Julesz. Binocular depth perception of computer-generated patterns. *Bell System Technical Journal*, 39(5):1125–1162, 1960. 15, 16, 68, 83, 124

[151] G. Sperling. Binocular vision: A physical and a neural theory. *The American Journal of Psychology*, 83(4):461–534, 12 1970. 15, 16, 17, 18, 139

[152] E. Hering, B. Bridgeman, and L. Stark. *The theory of binocular vision (1868)*. Springer, 1977. 15

[153] K. Koffka. *Principles of Gestalt psychology*, volume 44. Routledge, 1935.

[154] E. G. Boring. *Sensation and perception in the history of experimental psychology*. The Century psychology series. Irvington Publishers, New York, 1942.

[155] K. N. Ogle. *Researches in binocular vision.* WB Saunders, 1950.

[156] R. S. Woodworth and H. Schlosberg. *Experimental psychology*. Oxford and IBH Publishing, 1954. 15

[157] N. Chomsky. *Aspects of the theory of syntax*, volume 11. MIT Press, Cambridge, MA, 1965. 16

[158] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194(4262):283–7, Oct 1976. 16, 17, 18, 25, 67, 124, 139

[159] M. Carandini. From circuits to behavior: a bridge too far? *Nature Publishing Group*, 15(4):507–509, April 2012. 16

[160] B. L. Anderson. Can computational goals inform theories of vision? *Top Cogn Sci*, 7(2):274–86, Apr 2015. 16

[161] J. E. Mayhew and J. P. Frisby. Psychophysical and computational studies towards a theory of human stereopsis. *Artificial Intelligence*, 17(1):349 – 385, 1981. 16

[162] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proc R Soc Lond B Biol Sci*, 204(1156):301–28, May 1979. 16, 17, 18

[163] W. E. Grimson. A computer implementation of a theory of human stereo vision. *Philos Trans R Soc Lond B Biol Sci*, 292(1058):217–53, May 1981. 16, 18

[164] V. S. Ramachandran, V. M. Rao, and T. R. Vidyasagar. The role of contours in steropsis. *Nature*, 242(5397):412–4, Apr 1973. 16

[165] V. S. Ramachandran and J. I. Nelson. Global grouping overrides point-to-point disparities. *Perception*, 5(2):125–8, 1976. 16

[166] K. Prazdny. Detection of binocular disparities. *Biol Cybern*, 52(2):93–9, 1985. 16, 17

[167] J. I. Nelson. Globality and stereoscopic fusion in binocular vision. *J Theor Biol*, 49(1):1–88, Jan 1975. 16, 17, 18

[168] R. Szeliski and G. Hinton. Solving random-dot stereograms using the heat equation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'85)*, pages 284–288, San Francisco, June 1985. IEEE Computer Society Press. 17

[169] P. Dev. Perception of depth surfaces in random-dot stereograms : a neural model. *International Journal of Man-Machine Studies*, 7(4):511 – 528, 1975. 17, 18

[170] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. Pmf: A stereo correspondence algorithm using a disparity gradient limit. *Perception*, 14(4):449–470, 1985. 17

[171] P. Burt and B. Julesz. Modifications of the classical notion of panum's fusional area. *Perception*, 9(6):671–82, 1980. 17

[172] B. Gillam, S. Blackburn, and K. Nakayama. Stereopsis based on monocular gaps: metrical encoding of depth and slant without matching contours. *Vision Res*, 39(3):493–502, Feb 1999. 18

[173] P. N. Belhumeur and D. Mumford. A bayesian treatment of the stereo correspondence problem using half-occluded regions. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, pages 506–512. IEEE, 1992. 18

[174] P. N. Belhumeur. A bayesian approach to binocular steropsis. *International Journal of Computer Vision*, 19(3):237–260, 1996.

[175] B. L. Anderson and B. Julesz. A theoretical analysis of illusory contour formation in stereopsis. *Psychological Review*, 102(4):705, 1995. 18

[176] I. Tsirlin, L. M. Wilcox, and R. S. Allison. A computational theory of da vinci stereopsis. *Journal of vision*, 14(7), Jun 2014. 18, 34

[177] G. Maiello, M. Chessa, F. Solari, and P. J. Bex. Simulated disparity and peripheral blur interact during binocular fusion. *J Vis*, 14(8):13, 2014. 18

[178] C. Enroth-Cugell and J. G. Robson. The contrast sensitivity of retinal ganglion cells of the cat. *J Physiol*, 187(3):517–52, Dec 1966. 19, 20

[179] R. Shapley and S. Hochstein. Visual spatial summation in two classes of geniculate cells. *Nature*, 256(5516):411–3, Jul 1975. 19, 20

[180] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol*, 283:53–77, Oct 1978. 19, 47

[181] L. Maffei and A. Fiorentini. The visual cortex as a spatial frequency analyser. *Vision Res*, 13(7):1255–67, Jul 1973. 19

[182] M. Carandini, D. J. Heeger, and J. A. Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17:8621–8644, 1997. 19

[183] S. Hochstein and R. M. Shapley. Linear and nonlinear spatial subunits in y cat retinal ganglion cells. *J Physiol*, 262(2):265–84, Nov 1976. 20

[184] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst. Receptive field organization of complex cells in the cat's striate cortex. *J Physiol*, 283:79–99, Oct 1978. 20

[185] J. Lippert and H. Wagner. A threshold explains modulation of neural responses to opposite-contrast stereograms. *Neuroreport*, 12(15):3205–8, Oct 2001. 23

[186] J. C. A. Read, A. J. Parker, and B. G. Cumming. A simple model accounts for the response of disparity-tuned v1 neurons to anticorrelated images. *Vis Neurosci*, 19(6):735–53, 2002. 23, 29, 40, 55, 84

[187] J. C. A. Read and B. G. Cumming. Testing quantitative models of binocular disparity selectivity in primary visual cortex. *J. Neurophysiol.*, 90(5):2795–2817, November 2003. 23, 54, 55

[188] S. Tanabe and B. G. Cumming. Mechanisms underlying the transformation of disparity signals from v1 to v2 in the macaque. *J Neurosci*, 28(44):11304–14, Oct 2008. 23

[189] N. Qian and Y. Zhu. Physiological computation of binocular disparity. *Vision Res*, 37(13):1811–27, Jul 1997. 23, 26, 40, 55

[190] N. Qian. Computing stereo disparity and motion with known binocular cell properties. *Neural Comput.*, 6(3):390–404, May 1994. 23

[191] D. J. Fleet, H. Wagner, and D. J. Heeger. Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision research*, 36(12):1839–1857, June 1996. 23, 26, 40, 55, 60

[192] J. C. A. Read and B. G. Cumming. Sensors for impossible stimuli may solve the stereo correspondence problem. *Nat Neurosci*, 10(10):1322–8, Oct 2007. 23, 26, 39

[193] T. J. Preston, S. Li, Z. Kourtzi, and A. E. Welchman. Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *J. Neurosci.*, 28(44):11315–11327, October 2008. 24, 87, 89, 90, 100, 119, 121, 124, 125, 127

[194] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal Of Computer Vision*, 47(1-3):7–42, 2002. 25, 67

[195] B. G. Cumming and G. C. DeAngelis. The physiology of stereopsis. *Annual Review of Neuroscience*, 24:203–238, 2001. 25

[196] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. 26, 27, 94, 111

[197] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency Detection on Light Field. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2806–2813, 2014. 29, 45, 61

[198] J. Lippert and H. Wagner. A threshold explains modulation of neural responses to opposite-contrast stereograms. *Neuroreport*, 12(15):3205–3208, October 2001. 29

[199] J. Harris and A. J. Parker. Independent neural mechanisms for bright and dark information in binocular stereopsis. *Nature*, 374(6525):808–811, 1995. 32, 33, 64

[200] J. C. A. Read, X. A. Vaz, and I. Serrano-Pedraza. Independent mechanisms for bright and dark image features in a stereo correspondence task. *Journal of vision*, 11(12), 2011. 32, 33, 49

[201] B. L. Anderson and K. Nakayama. Toward a general theory of stereopsis: binocular matching, occluding contours, and fusion. *Psychological Review*, 101(3):414–445, July 1994. 34

[202] D. Brewster. XLIII.—On the Knowledge of Distance given by Binocular Vision. *Transactions of the Royal Society of Edinburgh*, 15(04):663–675, January 1844. 34

[203] S. P. Mckee, P. Verghese, A. Ma-Wyatt, and Y. Petrov. The wallpaper illusion explained. *Journal of vision*, 7(14):10.1–10, November 2007. 34

[204] H. B. Barlow. Summation and inhibition in the frog's retina. *The Journal of physiology*, 119(1):69–88, January 1953. 38

[205] H. B. Barlow. Possible principles underlying the transformations of sensory messages. In W. Rosenblith, editor, *Sensory Communication*, pages 217–234. MIT Press, Cambridge, MA, 1961. 38

[206] E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, 2001. 38

[207] Y. Karklin and M. S. Lewicki. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, 457(7225):83–86, January 2009. 38

[208] Z. Li and J. J. Atick. Efficient stereo coding in the multiscale representation*. *Network-Computation In Neural Systems*, 5:157–174, 1994. 38

[209] J. Burge and W. S. Geisler. Optimal disparity estimation in natural stereo images. *J Vis*, 14(2), 2014. 38, 84

[210] K. Okajima. Binocular disparity encoding cells generated through an Infomax based learning algorithm. *Neural networks : the official journal of the International Neural Network Society*, 17(7):953–962, September 2004. 38

[211] D. W. Hunter and P. B. Hibbard. Distribution of independent components of binocular natural images. *J Vis*, 15(13):6, 2015.

[212] D. W. Hunter and P. B. Hibbard. Ideal Binocular Disparity Detectors Learned Using Independent Subspace Analysis on Binocular Natural Image Pairs. *PLoS ONE*, 11(3):e0150117, 2016. 38

[213] A. Anzai, I. Ohzawa, and R. Freeman. Joint-encoding of motion and depth by visual cortical neurons: neural basis of the Pulfrich effect. *Nature Neuroscience*, 4(5):513–518, 2001. 40

[214] A. Nieder and H. Wagner. Hierarchical processing of horizontal disparity information in the visual forebrain of behaving owls. *J Neurosci*, 21(12):4514–22, Jun 2001. 41, 51, 85

[215] R. Blake and H. Wilson. Binocular vision. *Vision research*, 51(7):754–770, April 2011. 41

[216] R. Blake. Binocular rivarly and stereopsis revisited. In *From perception to consciousness*, pages 1–8. Oxford University Press, Oxford, March 2012. 41

[217] A. A. Muryy, R. W. Fleming, and A. E. Welchman. Key characteristics of specular stereo. *Journal of vision*, 14(14):14–14, December 2014. 41

[218] A. A. Muryy, R. W. Fleming, and A. E. Welchman. 'Proto-rivalry': how the binocular brain identifies gloss. *Proceedings Of The Royal Society B-Biological Sciences*, 283(1830):20160383, May 2016. 41

[219] A. M. van Loon, T. Knapen, H. S. Scholte, E. St John-Saaltink, T. H. Donner, and V. A. F. Lamme. GABA Shapes the Dynamics of Bistable Perception. *Current biology : CB*, 23(9):823–827, May 2013. 41

[220] C. Lunghi, U. E. Emir, M. C. Morrone, and H. Bridge. Short-Term Monocular Deprivation Alters GABA in the Adult Human Visual Cortex. *Current Biology*, 25(11):1496–1501, June 2015. 41

[221] J. W. Nadler, D. E. Angelaki, and G. C. Deangelis. A neural representation of depth from motion parallax in macaque visual cortex. *Nature*, 452(7187):642–645, April 2008. 42

[222] H. R. Kim, D. E. Angelaki, and G. C. Deangelis. A novel role for visual perspective cues in the neural computation of depth. *Nature Neuroscience*, 18(1):129–137, December 2014.

[223] M. L. Morgan, G. C. Deangelis, and D. E. Angelaki. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, 59(4):662–673, August 2008. 42

[224] M. P. Wellman and M. Henrion. Explaining 'explaining away'. *IEEE transactions on pattern analysis and machine intelligence*, 15(3):287–292, March 1993. 42

[225] H. R. Kim, X. Pitkow, D. E. Angelaki, and G. C. Deangelis. A simple approach to ignoring irrelevant variables by population decoding based on multisensory neurons. *Journal Of Neurophysiology*, 116(3):1449–1467, September 2016. 42

[226] R. Moreno-Bote, J. Beck, I. Kanitscheider, X. Pitkow, P. Latham, and A. Pouget. Information-limiting correlations. *Nat Neurosci*, 17(10):1410–7, Oct 2014. 44

[227] The Theano Development Team, R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, Y. Bengio, A. Bergeron, J. Bergstra, V. Bisson, J. Bleecher Snyder, N. Bouchard, N. Boulanger-Lewandowski, X. Bouthillier, A. de Brébisson, O. Breuleux, P.-L. Carrier, K. Cho, J. Chorowski, P. Christiano, T. Cooijmans, M.-A. Côté, M. Côté, A. Courville, Y. N. Dauphin, O. Delalleau, J. Demouth, G. Desjardins, S. Dieleman, L. Dinh, M. Ducoffe, V. Dumoulin, S. Ebrahimi Kahou, D. Erhan, Z. Fan, O. Firat, M. Germain, X. Glorot, I. Goodfellow, M. Graham, C. Gulcehre, P. Hamel, I. Harlouchet, J.-P. Heng, B. Hidasi, S. Honari, A. Jain, S. Jean, K. Jia, M. Korobov, V. Kulkarni, A. Lamb, P. Lamblin, E. Larsen, C. Laurent, S. Lee, S. Lefrancois, S. Lemieux, N. Léonard, Z. Lin, J. A. Livezey, C. Lorenz, J. Lowin, Q. Ma, P.-A. Manzagol, O. Mastropietro, R. T. McGibbon, R. Memisevic, B. van Merriënboer, V. Michalski, M. Mirza, A. Orlandi, C. Pal, R. Pascanu, M. Pezeshki, C. Raffel, D. Renshaw, M. Rocklin, A. Romero, M. Roth, P. Sadowski, J. Salvatier, F. Savard, J. Schlüter, J. Schulman, G. Schwartz, I. Vlad Serban, D. Serdyuk, S. Shabanian, É. Simon, S. Spieckermann, S. Ramana Subramanyam, J. Sygnowski, J. Tanguay, G. van Tulder, J. Turian, S. Urban, P. Vincent, F. Visin, H. de Vries, D. Warde-Farley, D. J. Webb, M. Willson, K. Xu, L. Xue, L. Yao, S. Zhang, and Y. Zhang. Theano: A Python framework for fast computation of mathematical expressions. *ArXiv e-prints*, May 2016. 46

[228] R. L. De Valois, D. G. Albrecht, and L. G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res*, 22(5):545–59, 1982. 50

[229] M. Jazayeri and J. A. Movshon. Optimal representation of sensory information by neural populations. *Nat Neurosci*, 9(5):690–6, May 2006. 57

[230] F. Longordo, M.-S. To, K. Ikeda, and G. J. Stuart. Sublinear integration underlies binocular processing in primary visual cortex. *Nature neuroscience*, 16(6):714–723, Jun 2013. 60, 66

[231] N. R. Goncalves and A. E. Welchman. "what not" detectors help the brain see in depth. *Curr Biol*, 27(10):1403–1412.e8, May 2017. 68, 80, 83, 84, 85

[232] D. H. Brainard. The psychophysics toolbox. *Spat Vis*, 10(4):433–6, 1997. 69

[233] D. G. Pelli. The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*, 10(4):437–42, 1997.

[234] M. Kleiner, D. Brainard, D. Pelli, A. Ingling, R. Murray, C. Broussard, et al. What's new in psychtoolbox-3. *Perception*, 36(14):1, 2007. 69

[235] A. B. Watson and D. G. Pelli. Quest: a bayesian adaptive psychometric method. *Percept Psychophys*, 33(2):113–20, Feb 1983. 70

[236] I. Fründ, N. V. Haenel, and F. A. Wichmann. Inference for psychometric functions in the presence of nonstationary behavior. *J Vis*, 11(6), May 2011. 72, 77

[237] J. Read. Early computational processing in binocular vision and depth perception. *Prog Biophys Mol Biol*, 87(1):77–108, Jan 2005. 73

[238] C. W. Tyler. Spatial organization of binocular disparity sensitivity. *Vision Res.*, 15(5):583–590, 1975. 76

[239] D. R. Badcock and C. M. Schor. Depth-increment detection function for individual spatial channels. *J Opt Soc Am A*, 2(7):1211–6, Jul 1985. 83, 117, 120

[240] S. B. Stevenson, L. K. Cormack, C. M. Schor, and C. W. Tyler. Disparity tuning in mechanisms of human stereopsis. *Vision Res*, 32(9):1685–94, Sep 1992. 76, 78, 83, 98, 112, 117, 120, 123

[241] N. R. Goncalves, H. Ban, R. M. Sánchez-Panchuelo, S. T. Francis, D. Schluppeck, and A. E. Welchman. 7 tesla fmri reveals systematic functional organization for binocular disparity in dorsal visual cortex. *J Neurosci*, 35(7):3056–72, Feb 2015. 78

[242] B. G. Breitmeyer, W. S. Hoar, D. Randall, and F. P. Conte. *Visual masking: An integrative approach*. Clarendon Press, 1984. 80

[243] H. v. Helmholtz. Physiological optics, vol. 3, trans. *Optical Society of America.*, 1909. 83

[244] S. Henriksen, B. G. Cumming, and J. C. A. Read. A single mechanism can account for human perception of depth in mixed correlation random dot stereograms. *PLOS Computational Biology*, 12(5):e1004906, May 2016. 84

[245] S. A. Chowdhury and G. C. DeAngelis. Fine discrimination training alters the causal contribution of macaque area mt to depth perception. *Neuron*, 60(2):367–77, Oct 2008. 84

[246] D. H. F. Chang, C. Mevorach, Z. Kourtzi, and A. E. Welchman. Training transfers the limits on perception from parietal to ventral cortex. *Curr Biol*, 24(20):2445–50, Oct 2014. 84

[247] P. Jaini and J. Burge. Linking normative models of natural tasks to descriptive models of neural response. *bioRxiv*, 2017. 84

[248] J. M. Harris, S. P. McKee, and H. S. Smallman. Fine-scale processing in human binocular stereopsis. *J Opt Soc Am A Opt Image Sci Vis*, 14(8):1673–83, Aug 1997. 85

[249] M. S. Banks, S. Gepshtein, and M. S. Landy. Why is spatial stereoresolution so low? *J Neurosci*, 24(9):2077–89, Mar 2004. 85

[250] H. Nienborg, H. Bridge, A. J. Parker, and B. G. Cumming. Receptive field size in v1 neurons limits acuity for perceiving disparity modulation. *J Neurosci*, 24(9):2065–76, Mar 2004. 85

[251] D. Y. Ts'o, M. Zarella, and G. Burkitt. Whither the hypercolumn? *J Physiol*, 587(Pt 12):2791–805, Jun 2009. 86

[252] A. W. Roe and D. Y. Ts'o. Visual topography in primate v2: multiple representation across functional stripes. *J Neurosci*, 15(5 Pt 2):3689–715, May 1995. 87, 121

[253] D. Y. Ts'o, A. W. Roe, and C. D. Gilbert. A hierarchy of the functional organization for color, form and disparity in primate visual area v2. *Vision Res*, 41(10-11):1333–49, 2001. 87

[254] W. van der Zwaag, S. Francis, K. Head, A. Peters, P. Gowland, P. Morris, and R. Bowtell. fmri at 1.5, 3 and 7 t: characterising bold signal changes. *Neuroimage*, 47(4):1425–34, Oct 2009. 87, 100

[255] K. Cheng, R. A. Waggoner, and K. Tanaka. Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron*, 32(2):359–74, Oct 2001. 87, 105, 121

[256] E. Yacoub, N. Harel, and K. Ugurbil. High-field fMRI unveils orientation columns in humans. *PNAS*, 105(30):10607–10612, July 2008. 87, 105, 122

[257] J. Zimmermann, R. Goebel, F. De Martino, P.-F. van de Moortele, D. Feinberg, G. Adriany, D. Chaimow, A. Shmuel, K. Ugurbil, and E. Yacoub. Mapping the Organization of Axis of Motion Selective Features in Human Area MT Using High-Field fMRI. *PLoS ONE*, 6(12):e28716, December 2011. 87, 121, 122, 128

[258] A. M. Dale, B. Fischl, and M. I. Sereno. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194, February 1999. 89

[259] B. Fischl, M. I. Sereno, and A. M. Dale. Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9(2):195–207, February 1999. 89

[260] J. S. Gati, R. S. Menon, K. Uğurbil, and B. K. Rutt. Experimental determination of the bold field strength dependence in vessels and tissue. *Magnetic Resonance in Medicine*, 38(2):296–302, 1997. 90, 122, 136

[261] S. Ogawa, R. S. Menon, S. G. Kim, and K. Ugurbil. On the characteristics of functional magnetic resonance imaging of the brain. *Annu Rev Biophys Biomol Struct*, 27:447–74, 1998.

[262] K. Uğurbil, G. Adriany, P. Andersen, W. Chen, M. Garwood, R. Gruetter, P.-G. Henry, S.-G. Kim, H. Lieu, I. Tkac, T. Vaughan, P.-F. Van De Moortele, E. Yacoub, and X.-H. Zhu. Ultrahigh field magnetic resonance imaging and spectroscopy. *Magn Reson Imaging*, 21(10):1263–81, Dec 2003. 90, 122, 136

[263] R. M. Sanchez-Panchuelo, J. Besle, A. Beckett, R. Bowtell, D. Schluppeck, and S. Francis. Within-digit functional parcellation of Brodmann areas of the human primary somatosensory cortex using functional magnetic resonance imaging at 7 tesla. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(45):15815–15822, November 2012. 90, 122

[264] J. R. Polimeni, B. Fischl, D. N. Greve, and L. L. Wald. Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *NeuroImage*, 52(4):1334–1346, October 2010. 90, 100, 122, 129

[265] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May 2011. 90

[266] A. S. Rojer and E. L. Schwartz. Cat and monkey cortical columnar patterns modeled by bandpass-filtered 2d white noise. *Biol. Cybern.*, 62(5):381–391, 1990. 92, 102, 106

[267] G. M. Boynton. Imaging orientation selectivity: decoding conscious perception in v1. *Nat. Neurosci.*, 8(5):541–542, May 2005. 92

[268] C. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949. 94, 111

[269] G. C. DeAngelis and T. Uka. Coding of horizontal disparity and velocity by mt neurons in the alert macaque. *J. Neurophysiol.*, 89(2):1094–1111, February 2003. 95, 97

[270] M. B. Hoffmann, J. Stadler, M. Kanowski, and O. Speck. Retinotopic mapping of the human visual cortex at a magnetic field strength of 7 t. *Clinical Neurophysiology*, 120(1):108–116, 2016/04/23 2009. 100

[271] H. Ban, T. J. Preston, A. Meeson, and A. E. Welchman. The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat Neurosci*, 15(4):636–643, February 2012. 104, 119

[272] A. P. Murphy, H. Ban, and A. E. Welchman. Integration of texture and disparity cues to surface slant in dorsal visual cortex. *J Neurophysiol*, 110(1):190–203, Jul 2013.

[273] D. Dövencioğlu, H. Ban, A. J. Schofield, and A. E. Welchman. Perceptual integration for qualitatively different 3-d cues in the human brain. *J Cogn Neurosci*, 25(9):1527–41, Sep 2013. 119

[274] B. R. Cottereau, S. P. McKee, J. M. Ales, and A. M. Norcia. Disparity-tuned population responses from human visual cortex. *J Neurosci*, 31(3):954–65, Jan 2011. 121

[275] F. De Martino, J. Zimmermann, L. Muckli, K. Ugurbil, E. Yacoub, and R. Goebel. Cortical depth dependent functional responses in humans at 7t: improved specificity with 3d grase. *PLoS One*, 8(3):e60514, 2013. 122

[276] A. Shmuel, E. Yacoub, D. Chaimow, N. K. Logothetis, and K. Ugurbil. Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. *NeuroImage*, 35(2):539–552, April 2007. 122

[277] B. G. Cumming and A. J. Parker. Binocular neurons in v1 of awake monkeys are selective for absolute, not relative, disparity. *J Neurosci*, 19(13):5602–18, Jul 1999. 124

[278] K. Krug and A. J. Parker. Neurons in dorsal visual area v5/mt signal relative disparity. *J Neurosci*, 31(49):17892–904, Dec 2011. 124

[279] M. W. Self, T. van Kerkoerle, H. Supèr, and P. R. Roelfsema. Distinct roles of the cortical layers of area v1 in figure-ground segregation. *Curr Biol*, Oct 2013. 125

[280] T. van Kerkoerle, M. W. Self, and P. R. Roelfsema. Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat Commun*, 8:13804, Jan 2017. 125

[281] P. Kok, L. J. Bains, T. van Mourik, D. G. Norris, and F. P. de Lange. Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Curr Biol*, 26(3):371–6, Feb 2016. 125

[282] L. Muckli, F. De Martino, L. Vizioli, L. S. Petro, F. W. Smith, K. Ugurbil, R. Goebel, and E. Yacoub. Contextual feedback to superficial layers of v1. *Curr Biol*, 25(20):2690–5, Oct 2015. 125

[283] S. E. Jones, B. R. Buchbinder, and I. Aharon. Three-dimensional mapping of cortical thickness using laplace's equation. *Hum Brain Mapp*, 11(1):12–32, Sep 2000. 128, 133

[284] F. De Martino, J. Zimmermann, L. Muckli, K. Ugurbil, E. Yacoub, and R. Goebel. Cortical depth dependent functional responses in humans at 7t: Improved specificity with 3d grase. *PLoS ONE*, 8(3):e60514, 03 2013. 129, 131, 136, 137

[285] P. J. Koopmans, M. Barth, and D. G. Norris. Layer-specific BOLD activation in human V1. *Human Brain Mapping*, 31(9):1297–1304, August 2010. 129

[286] J. C. W. Siero, D. Hermes, H. Hoogduin, P. R. Luijten, N. F. Ramsey, and N. Petridou. Bold matches neuronal activity at the mm scale: a combined 7t fmri and ecog study in human sensorimotor cortex. *Neuroimage*, 101:177–84, Nov 2014. 136

[287] H. Duvernoy, S. Delon, and J. Vannson. Cortical blood vessels of the human brain. *Brain Research Bulletin*, 7(5):519 – 579, 1981. 136

[288] D. Feinberg, N. Harel, S. Ramanna, K. Ugurbil, and E. Yacoub. Sub-millimeter single-shot 3d grase with inner volume selection for t2 weighted fmri applications at 7 tesla. *Magn Reson Med*, 2008. 136

[289] M. D. Waehnert, J. Dinse, M. Weiss, M. N. Streicher, P. Waehnert, S. Geyer, R. Turner, and P.-L. Bazin. Anatomically motivated modeling of cortical laminae. *Neuroimage*, Apr 2013. 136

[290] L. Huber, J. Goense, A. J. Kennerley, R. Trampel, M. Guidi, E. Reimer, D. Ivanov, N. Neef, C. J. Gauthier, R. Turner, and H. E. Möller. Cortical lamina-dependent blood volume changes in human brain at 7 t. *Neuroimage*, 107:23–33, Feb 2015. 137

[291] D. G. Stork. *Hal's Legacy: 2001's Computer As Dream and Reality*. MIT Press, Cambridge, MA, USA, 1996. 140

[292] M. A. Smith, W. Bair, and J. A. Movshon. Dynamics of suppression in macaque primary visual cortex. *J Neurosci*, 26(18):4826–34, May 2006. 141

[293] T. Uka and G. C. DeAngelis. Contribution of area mt to stereoscopic depth perception: Choice-related response modulations reflect task strategy. *Neuron*, 42(2):297–310, April 2004. 142

[294] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–82, Jan 2011. 143

[295] Y. Trotter, S. Celebrini, B. Stricanne, S. Thorpe, and M. Imbert. Modulation of neural stereoscopic processing in primate area v1 by the viewing distance. *Science*, 257(5074):1279–81, Aug 1992. 144