The Human Brain in Depth: How We See in 3D

Andrew E. Welchman

Department of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom;
email: aew69@cam.ac.uk

Keywords

human brain imaging, stereopsis, depth cues, cue fusion, binocular vision,
sensory integration

Abstract

Human perception is remarkably flexible: We experience vivid three-
dimensional (3D) structure under diverse conditions, from the seemingly
random magic-eye stereograms to the aesthetically beautiful, but obviously
flat, canvases of the Old Masters. How does the brain achieve this apparently
effortless robustness? Using brain imaging we are beginning to discover how
different parts of the visual cortex support 3D perception by tracing different
computations in the dorsal and ventral pathways. This review concentrates
on studies of binocular disparity and its combination with other depth cues.
This work suggests that the dorsal visual cortex is strongly engaged by 3D
information and is involved in integrating signals to represent the structure
of viewed surfaces. The ventral cortex may store representations of object
configurations and the features required for task performance. These dif-
ferences can be broadly understood in terms of the different computational
demands of reducing estimator variance versus increasing the separation
between exemplars.

INTRODUCTION

We inhabit a world of illuminated physical objects. The visual system is tasked with appreciating the properties of these objects from a distance so that we might move among them and engage with them. This process underlies our skills in navigation, our aesthetic appreciation of quality, and our selection of appropriate foods and conspecifics. Traditional approaches stemming from engineering have encouraged us to compartmentalize these skills [e.g., visual navigation, motion, three-dimensional (3D) shape, object recognition, color, face perception] and break down the inputs into manageable chunks (e.g., illuminant cues, chromaticity signals, shape from shading). Here, I use this logic to review the mechanisms that support our perception of 3D structure. However, this is a convenient fiction. All vision is three-dimensional: It depends on complex interactions between objects, illuminants, and viewers that occupy different locations in space. My focus is on the human brain; although I link to work in animal models, extensive reviews are provided elsewhere (Cumming & DeAngelis 2001, Gonzalez & Perez 1998, Orban 2011, Parker 2007).

TAKING THE BRAIN TO ANOTHER DIMENSION

Our sensory apparatus sample the surrounding environment by projecting it into a lower-dimensional measurement space that provides a set of descriptors sufficient to promote our survival. So we hear vibration energy over a limited frequency range and sample electromagnetic radiation using detectors with limited spectral sensitivity (e.g., cone bandwidths of approximately 100 nm). The classic starting point for discussing 3D vision is the potential difficulties that result from dimensionality reduction: How can we estimate a 3D world using a 2D projection onto the retina? This problem is particularly acute when we think about scenes composed of isolated points and thin lines (**Figure 1a**).

Figure 1

(*a*) The binocular correspondence problem. Given two retinal images, how does the brain match elements between the eyes? Nine potential matches are illustrated plus two ghost half matches. Filled yellow circles represent a stable match. The reader might also experience (by fusing the boxes below the eyes) four targets (*filled orange circle*s), two of which are ghost matches. (*b*) Illustrations of some of the different types of information used by the brain to gain an impression of depth.

Some further thought might suggest that the answer to this question is straightforward. For instance, we have two eyes and, thus, can triangulate the true depth of an object from the subtle differences in positions registered on the two retinae (binocular disparity). Moreover, we have access to a range of visual signals related to depth: These cues can, in isolation or combination, give rise to an impression of 3D structure (**Figure 1***b*). Perhaps the answer to the challenge of perceiving three dimensions is, therefore, trivial: A rich natural world gives us more than enough information to resolve ambiguity and explain our impression of depth. Yet it is not clear how trivial such a solution would be to the neurons of the early visual cortex. Their responses resemble a series of points and lines, posing a challenge in understanding how the brain translates from locally ambiguous signals to extract the properties of extensive surfaces and objects (Wallach 1935). Thus, although at the system level we can think about the range of signals that come together to effect our impression of depth, it is clear there is some work to do in understanding the biological computations that are responsible for this outcome.

Let us start trying to unpick the gray box between our ears by thinking about the nature of the inputs and the outputs. The range of our natural viewing experiences suggests we can infer depth in diverse environments. The most remarkable observations are, perhaps, that we perceive depth under impoverished conditions: a collection of seemingly random dots (Julesz 1971, Wallach & O'Connell 1953) or an objectively flat painting viewed away from the artist's center of perspective projection (Pirenne 1970). This encourages us to think about multiple routes in and our semantic classifications that carve visual inputs into different cues (**Figure 1***b*). By the same token, considering the functional uses to which 3D information is put, it is clear that there are multiple routes out. For instance, disparity can be used to infer 3D structure, break camouflage (Babington-Smith 1958), estimate surface reflectance (Blake & Bülthoff 1990), or control grasping (Watt & Bradshaw 2002).

Having multiple routes in and multiple routes out has the potential to complicate our understanding of the neural mechanisms responsible for depth. Imagine measuring the neural response to a disparity-defined 3D shape. What would this response reveal about the underlying cortical process? Does the response correspond to low-level disparity statistics? Is it an estimate of depth from the disparity cue (but not others)? Or is it a full estimate of shape based on integrating the available sensory signals with prior knowledge? Moreover, what is the purpose of the representation? Does it simply act as a cue to segmentation, or support judgments of shape, or provide a generalized depth map? We need to consider these issues when interpreting observations of neuronal responses.

WITH MODULARITY IN MIND

To ground our thinking in the face of these complications, it is useful to have a conceptual framework of potential processing architectures. Clark & Yuille (1990) provided a fundamentally important theoretical discussion of the ways in which cues might interact within an information processing system. They discussed a continuum of organizations, ranging from processing within separable modules that are linearly combined (weakly coupled fusion) to complex (nonlinear) interactions between cues (strongly coupled fusion). Under the weak scenario, cues are processed independently in separable modules to yield an impression of depth. So, if looking at an image depicting depth from disparity and shading (**Figure 2***a*), separate modules would process the two cues (disparity- versus shading units). Our impression of depth would come from fusing the outputs of these units (**Figure 2***b*). Alternatively, disparity and shading may interact with each other's processing to mutually constrain the interpretation of each: For example, certain interpretations of the shading cue could be ruled out based on the information provided by disparity (Bülthoff &
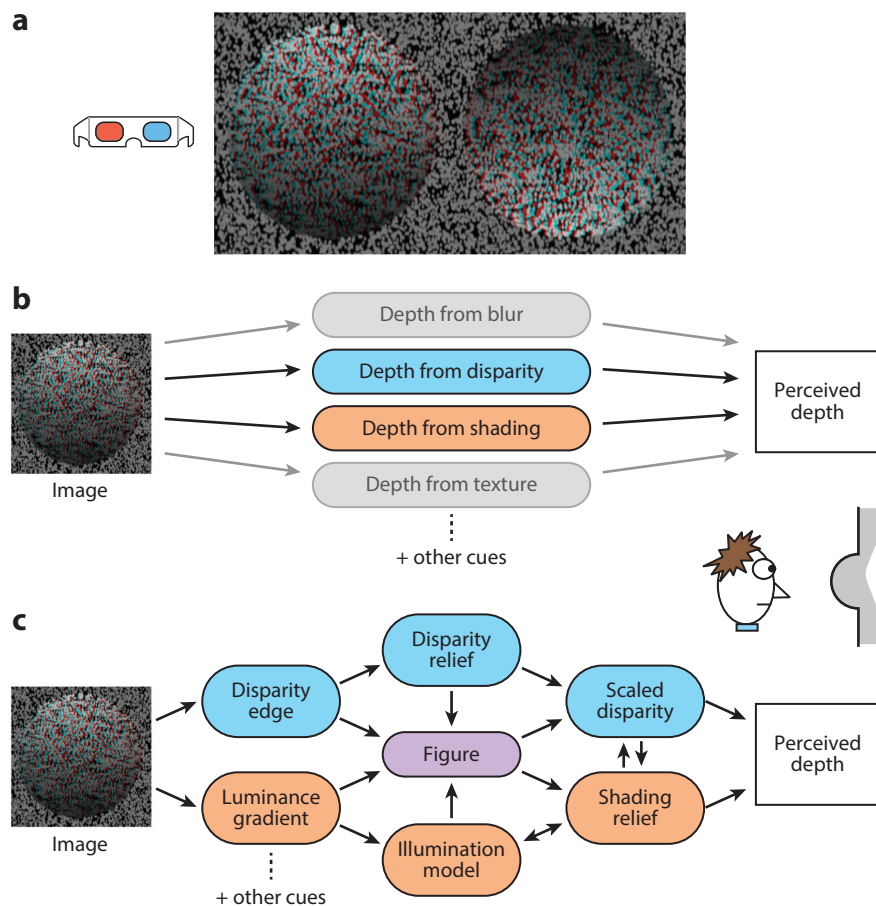
a

b

Depth from blur

Depth from disparity

Depth from shading

Depth from texture

Image

+ other cues

Perceived depth

c

Disparity edge

Disparity relief

Figure

Luminance gradient

Illumination model

Scaled disparity

Shading relief

Perceived depth

Image

+ other cues

Figure 2

(*a*) A stereogram depicting a dimpled surface defined by disparity and shading cues. This image, and stereograms in subsequent figures, are designed for viewing through red-cyan anaglyph glasses: red filter over left eye. (*b*) Weak fusion scenario for the modular processing of depth from each cue. The gray boxes indicate other cues that may also be engaged by the stimuli. (*c*) Strong fusion scenario that involves extensive common processing and interactions between cues to determine the perceived shape. This architecture is schematic and not intended to provide a complete description of how the brain estimates depth from these cues. Information from other cues may also contribute but is not illustrated for simplicity.

Mallot 1988). Under this scenario (**Figure 2*c***), we effectively lose the concept of separable modules because the interactions between the specific signals provided by each cue become so extensive that their processing cannot be said to be independent. This makes predicting the outcome in untested situations (e.g., different curvatures or surface reliefs) difficult. Pragmatically, such an architecture is unattractive: The ability to predict behavior in novel situations is a key element of a linear systems approach.

Based on psychophysical evidence (Dosher et al. 1986; Johnston et al. 1993, 1994) and considerations of the quantitative differences between cues, Landy and colleagues (1995) made a convincing case for a modified version of the weak fusion framework. They advocated a modular architecture in which cues are processed quasi-independently, with limited interactions, to allow

signals to be brought into coregistration (cues are "promoted" into comparable measurement scales and cross-calibrate one another). This view has been strongly endorsed by psychophysical tests of integration both within and between sensory modalities. In particular, perceptual judgments in combined cue situations (e.g., shape from disparity and texture, or from disparity and haptics) can be predicted from judgments based on cue attributes measured separately (Ernst & Banks 2002, Hillis et al. 2002, Jacobs 1999, Knill & Saunders 2003); that is, depth(disparity + haptics) is predicted from depth(disparity) and depth(haptics). Remarkably, these predictions relate not only to the magnitude of the estimated shape (weighted linear combination), such as the perceived slant of a surface, but also to the precision with which participants can make their judgments (maximum likelihood estimation), such as the participants' variability in judging slant.

Despite this strong empirical case for modularity and its computational advantages (Marr 1976), it is not immediately clear how this translates to the human brain. In particular, modular specialization necessitates considerable redundancy at higher levels of the visual system, whereby similar operations are performed in parallel. Empirically, there is little evidence of cortical areas wholly specialized for the processing of a particular depth cue. Moreover, it is perhaps presumptuous to suppose that our linguistic labels of depth cues correspond to the information architecture employed by the brain (despite the labels' intuitive appeal). To consider some specific examples, neurons sensitive to disparity are found throughout the visual and parietal cortices (Orban et al. 2006, Parker 2007), suggesting widespread use of binocular signals. Tests of shading cues initially suggested a prescribed locus in the ventral stream (Georgieva et al. 2008). However, subsequent work that related brain activity to the perception of shape from shading indicated the importance of responses in the dorsal visual areas and parietal cortex (Gerardin et al. 2010), and tests of patients indicated that areas outside the ventral cortex are involved (Gillebert et al. 2015). Thus, any modules for depth signals within the human brain are more plausibly vertical (i.e., streams across the cortical hierarchy) (Ponce et al. 2008, Ts'o et al. 2001), with a degree of specialization at the level of individual cues within particular areas (or subareas) of the cortex (Chen et al. 2008).

USING IMAGING TO STUDY 3D VISION

Our understanding of the neural circuitry that underlies human 3D vision has improved dramatically during the past 20 years through the maturing of noninvasive imaging protocols. These techniques [principally functional magnetic resonance imaging (fMRI) and electroencephalography (EEG)] measure responses to visual stimulation aggregated over large numbers of neurons, meaning that spatial resolution is limited. Moreover, the signal-to-noise ratio is low relative to recordings of single neurons, so that measurements typically require extensive averaging across stimulus repetitions. Coupled with the slow temporal dynamics of the BOLD (blood oxygen level–dependent) signal used in fMRI, this has a rate-limiting effect on the number of stimulus conditions that can be recorded. Nevertheless, these methods offer wide coverage of cortical activity. For instance, **Figure 3** provides an overview of visual areas localized through the retinotopic specificity of fMRI responses (e.g., early visual areas where the position of retinal stimulation is mapped out in a systematic way across the cortical surface) and through the use of functional contrasts (e.g., the human motion complex, hMT+/V5, is localized by contrasting moving and static patterns). Moreover, by integrating fMRI measurements with psychophysics, we have the facility to relate visual processing to its functional use. To understand the application of these methods to 3D vision, it is worth spending a few moments thinking about the logic and potential pitfalls of the approach. (I focus on fMRI, although many issues are common to other techniques, such as EEG).
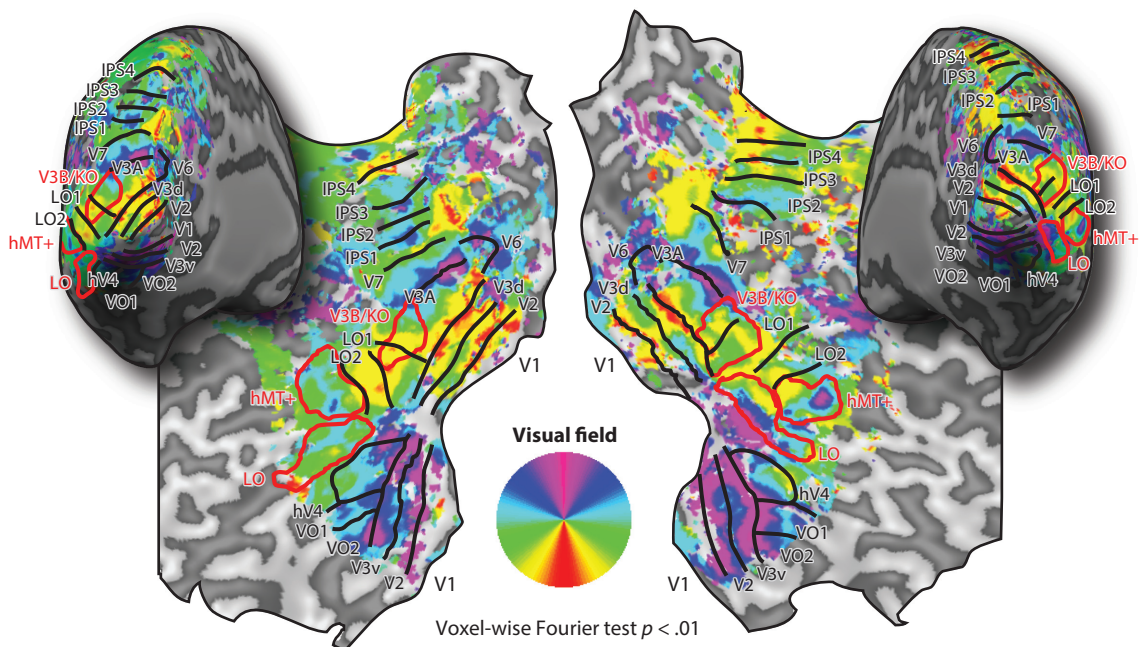
Visual field

Voxel-wise Fourier test *p* < .01

Figure 3

An illustration of visual areas within the human brain. The flat maps show the demarcation of different regions of interest in the left and right hemispheres of a single participant's brain. The activity pattern shows responses to a visual field mapping stimulus defined by a depth checkerboard. The V3B/KO (kinetic occipital) region is shown in relation to the lateral occipital (LO)1 and LO2 designations. The identification of LO1 and LO2 in this brain is difficult using polar and eccentricity mapping data. Figure created by Dr. Hiroshi Ban.

Functional imaging measurements are rarely absolute, but rather provide information about changes in brain activity. In designing an experiment, we therefore typically require a contrast between different stimulus conditions. A basic starting point is to establish that an area responds to a visual signal (e.g., stimulus on versus stimulus off). However, asking more detailed questions about processing requires more sophisticated contrasts.

Consider the investigation of whether an area encodes 3D structure from a cue such as shading. The choice of test stimulus might be obvious (a shading pattern that evokes a depth percept); however, it is often difficult to select an appropriate null stimulus. Ideally, the experimenter would choose a contrasting stimulus that has the same image features but does not give rise to a depth percept. In practice, this is difficult to achieve, so the choice of null stimulus will reflect our best guess at a control for a particular stage of visual processing (Georgieva et al. 2008). These choices can critically affect the inferences we draw about the locus of activity underlying the processing of a particular signal (**Figure 4a**). Moreover, if the low-level image features differ between depth and nondepth stimuli, it can be difficult to determine whether an fMRI response relates to simple, low-level features or the depth percept. Yet such simple, low-level feature differences are presumably responsible for determining whether a 3D structure is perceived (Fleming et al. 2011). Thus, it can be problematic to draw a distinction between the representation of low-level features versus depth structure. One approach is to randomize the images (e.g., type of texture or reflectance map) as much as possible and look for consistency in fMRI responses to a given 3D shape despite dramatic changes in the visual input. This approach can be successful, provided that images that appear radically different do not share the same underlying structure (Fleming et al. 2004).
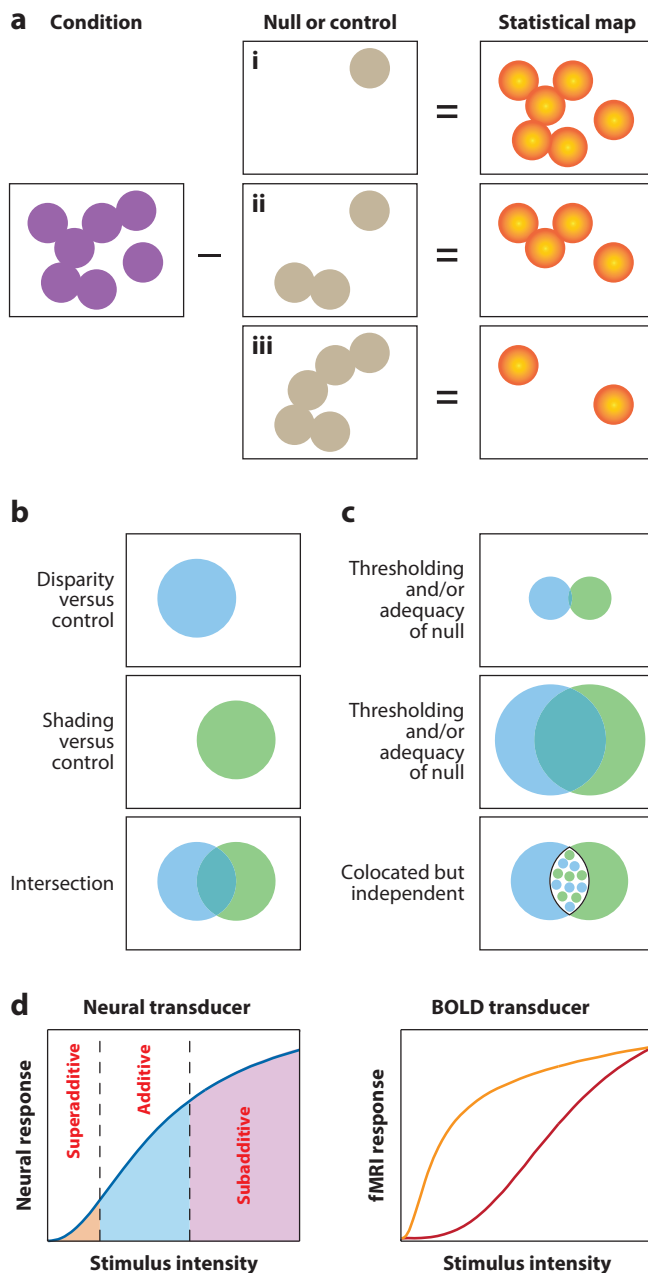
Figure 4

Illustrations of the logic of traditional brain imaging. (*a*) The logic of contrasts. The box depicts the brain, with a stimulus condition evoking a pattern of activity. A contrasting control or null stimulus is shown; activity is subtracted from that for the stimulus condition, yielding a map of brain activity. The choice of null stimulus (*i*, *ii*, or *iii*) affects the localized brain activity. (*b*) Test for multicue processing based on overlapping activity. (*c*) Potential interpretive difficulties with *b*. (*d*) Illustration of the transducer functions that relate stimulus intensity to evoked activity. Different regions of the curve have different properties for tests of additivity. The form of the BOLD transducer function is almost certainly different than for spiking neural activity, and may differ between depth cues (*red* versus *orange curves*). Abbreviations: BOLD, blood oxygen level–dependent; fMRI, functional magnetic resonance imaging.

If we now turn to the question of testing for depth responses from disparity, the choice of the null stimulus may seem simpler. In particular, a common logic is to contrast a "3D" condition (i.e., one in which the experimenter depicts binocular disparities) with a "2D" or "no disparity" condition (i.e., one in which the experimenter has not added binocular disparity to the display). This thinking is questionable for two reasons. First, technically it ignores the basic geometry of binocular vision. In particular, the locus of zero disparities (i.e., the horopter) depends of the fixation distance of the participant (Ogle 1932). If an image depicts a stimulus in the plane of the presentation screen (i.e., the "2D" viewing situation), then it is quite likely that there are considerable disparity variations in the stimulus because of the deviation of the display screen from the horopter. (These deviations can become quite substantial at short viewing distances or with stimuli that extend into the periphery more than a few degrees of visual angle, or both). Thus, such "2D" conditions will often provide an elaborate pattern of nonzero binocular disparities and may, therefore, reveal processes associated with the complex case of identifying that a presented surface is flat, rather than targeting low-level disparities (as is typically conceived in the experimental design). Second, and more importantly, there is good reason to believe that near-zero binocular disparities constitute the majority of our visual diet (Burge & Geisler 2014, Hibbard 2007). Single-unit measures indicate good selectivity to stimuli at, or very close to, zero (for summaries see Cumming & DeAngelis 2001), and psychophysics (Badcock & Schor 1985, Stevenson et al. 1992) suggests that humans are most sensitive to disparity information around zero. So it is unclear whether a zero disparity stimulus provides a good choice of null stimulus.
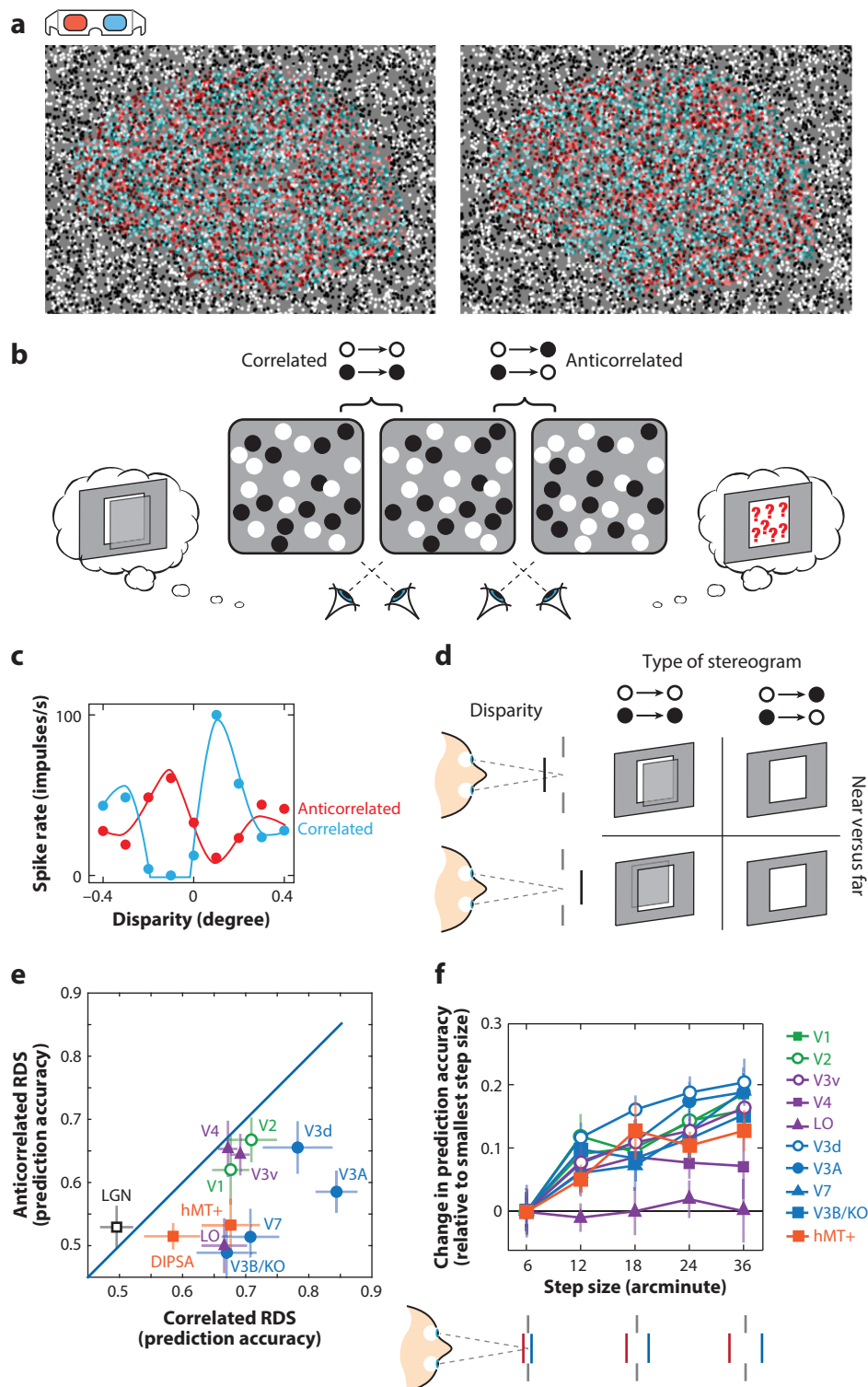
This is not to imply that the endeavor is hopeless: Manipulations of disparity content have provided important insight into the cortical landscape of neural responses to stimuli depicting depth (Durand et al. 2009, Georgieva et al. 2008, Tsao et al. 2003). However, it should be clear that choosing contrasting stimulus conditions is rarely simple, can affect the inferences we draw from contrasts, and that, in understanding the processing within a target region of the cortex, we need to think clearly about the information available. Parametric approaches, in which the investigator presents a systematic range of depth stimuli (Backus et al. 2001, Norcia et al. 1985, Regan & Spekreijse 1970), offer the potential to map changes in brain activity in terms of stimulus parameters. Moreover, combining different measures (e.g., fMRI with electrophysiology, behavioral judgments, or plausible computational processes) offers the potential to link neural and perceptual states (Welchman & Kourtzi 2013).

THE NEURAL BASIS OF DEPTH FROM BINOCULAR DISPARITY

Our understanding of stereoscopic processing in the primate brain owes much to Julesz's (1971) popularization of random dot stereograms (RDSs; **Figure 5**). At a conceptual level, these stimuli

---→

Figure 5

(*a*) Random dot stereograms (RDSs) depicting a disparity-defined object (a lateral view of the brain) in correlated and anticorrelated stimuli. When wearing red-cyan anaglyph glasses to view the stereogram (red filter over left eye), one should see coherent depth for the stimulus on the left but not on the right. (*b*) Construction of correlated and anticorrelated RDSs, and the perceptual result. (*c*) V1 neuron's response to RDS patterns as measured by Cumming & Parker (1997) (adapted with permission from Macmillan Publishers Ltd, *Nature*, 389, pp. 280–83, copyright 1997). (*d*) Logic of Preston et al.'s (2008) study using near versus far decoding with correlated and anticorrelated RDSs. (*e*,*f*) Results from Preston et al. (2008). (*e*) Decoding accuracies across visual hierarchy. Deviations from the diagonal show a preference for correlated RDSs. (*f*) Decoding accuracy as a function of the disparity difference between the near versus far stimuli (i.e., step size). Decoding accuracies are normalized across regions of interest. In particular, the decoding performance for a step size of 6 was subtracted from the decoding performance with the remaining disparity steps.

a

b

Correlated   ○ → ○   ○ → ●   Anticorrelated
             ● → ●   ● → ○

c

Spike rate (impulses/s)

Anticorrelated
Correlated

Disparity (degree)

d

Type of stereogram

○ → ○   ○ → ●
● → ●   ● → ○

Disparity

Near versus far

e

Anticorrelated RDS (prediction accuracy)

LGN

V4   V2   V3d

V3v
V1   hMT+   V3A
LO   V7
DIPSA   V3B/KO

Correlated RDS (prediction accuracy)

f

Change in prediction accuracy (relative to smallest step size)

■ V1
○ V2
○ V3v
■ V4
▲ LO
○ V3d
● V3A
▲ V7
■ V3B/KO
■ hMT+

Step size (arcminute)

highlighted that disparity could be divorced from figural representation, suggesting a modular architecture for depth. At an empirical level, the stimuli provided a powerful tool for identifying neural representations that are unambiguously binocular. In particular, manipulating the disparity of a bar stimulus might change neural responses, but it would be difficult to be certain that this was due to disparity rather than changes in the position of the bar in a single eye (Hubel & Wiesel 1970). RDS patterns enable us to avoid this potential covariate (Poggio et al. 1988), isolating changes at a cyclopean level of processing that is at, or after, binocular combination.

To review the processing of depth from disparity in the human, it is useful to think about discrete computational stages. (Neurobiology is undoubtedly less discrete.) First, the visual system needs to establish correspondence between the two eyes' views, coregistering image features so that disparity can be extracted. This correspondence problem appears particularly challenging when images comprise many self-similar elements (Julesz 1971). Second, local estimates of disparity need to be integrated and interpolated across space to form descriptors of spatially extensive features, such as edges, boundaries, and surfaces. Third, disparity estimates need to be related to real world properties through some type of scaling operation. Finally, disparity representations may be put to different purposes, so understanding a neuronal response necessitates relating it to performance on different types of task.

How Does the Brain Extract Disparities?

Given images presented to the two eyes, the first computational challenge is to match the inputs through a combination of vergence eye movements and neural coregistration (Marr & Poggio 1979). This process can be thought of as constrained local correlation between the left and right eyes' images, where disparity is defined by the spatial offset between correlated image items (Marr & Poggio 1976). The binocular energy model of Ohzawa and colleagues (1990) and its refinements (Fleet et al. 1996, Haefner & Cumming 2008, Read & Cumming 2007) provide, to a first approximation, a biological implementation of such correlation-based disparity detectors. Empirical measurements of V1 neurons suggest they are rather like binocular energy units: They respond selectively to disparities depicted in RDSs (e.g., **Figure 5a,b**) where stimuli are designed such that white dots in the left eye match white dots in the right eye (correlated RDS). Moreover, following the predictions of the energy model, the response profile of V1 neurons inverts (**Figure 5c**) if anticorrelated RDS stimuli are presented (a white dot in the left eye matches a black dot in the right eye; **Figure 5a,b**) (Cumming & Parker 1997). This is intriguing, as it suggests a marked dissociation between the responses of V1 neurons and the perceptual impression of depth. In particular, the disparity information contained in anticorrelated patterns does not support stereoscopic judgments (Cogan et al. 1993), suggesting that V1 neurons extract disparities that are not constrained by the global stereoscopic interpretation.

Testing responses evoked by correlated and anticorrelated RDSs provides a convenient means of separating neural responses to disparity per se as opposed to the perceptual experience of depth evoked by disparity, that is, disparity versus stereopsis. Bridge & Parker (2007) provided the first fMRI measurements with anticorrelated stimuli, finding an overall reduction in fMRI responses that was more marked in higher areas of the visual cortex. Preston et al. (2008) used correlated and anticorrelated RDSs to investigate a basic property of disparity encoding: whether a target is positioned near (crossed disparity) or far (uncrossed disparity) with respect to the fixation point. They used a decoding algorithm to quantify the discriminability of fMRI responses to near versus far depth positions, and drew a contrast between performance with correlated and anticorrelated RDS stimuli (**Figure 5d**). They showed that fMRI responses in early visual areas (V1, V2, V3v, V4, V3d) supported the decoding of disparity signals in anticorrelated patterns, even though

participants had no perception of the depth configuration (**Figure 5***e*). By contrast, in higher portions of the dorsal [V3A, V3B/KO (kinetic occipital), V7] and ventral [lateral occipital (LO)] visual streams, decoding performance was markedly higher for the correlated stimuli, suggesting disparity representations consistent with the stereoscopic experience of the participants.

In some respects Preston et al.'s (2008) results were confirmatory: Single-unit responses to anticorrelated RDSs are attenuated in macaque V4 (Tanabe et al. 2004) and absent in the inferior temporal (IT) cortex (Janssen et al. 2003). IT responses may be homologous to those in LO, and recordings from the dorsal portion of macaque V4 may correspond to activity in human V3B/KO. Nevertheless, by providing complete coverage of the visual cortex, the results raised a puzzle in that intermediate ventral areas (V3v, V4) supported robust decoding of anticorrelated patterns in marked contrast to higher ventral area LO. Therefore, the question arises of where the preference for correlated features in area LO originates. In particular, fMRI responses to anticorrelated stimuli do not appear to be progressively attenuated as information travels through intermediate ventral areas (V3v, V4) to terminate in LO. Rather, attenuation in the response to anticorrelated stimuli seems to be a property of the dorsal stream, suggesting that responses in LO may derive from activity in dorsal areas. This notion is necessarily speculative because fMRI does not reveal the nature of the intracortical interactions that give rise to responses in different cortical areas.

How Are Disparity Representations Organized?

These results suggest disparity representations consistent with stereoscopic perception in both dorsal and ventral visual areas. However, further tests by Preston et al. (2008) suggested that the form of the representations differed between pathways. In particular, by parametrically varying disparity, they found that dorsal areas (notably V3A) produced highly specific fMRI responses. By contrast, responses in LO were much less specific to the viewed disparity [e.g., the response to a disparity of 9 arcminutes (arcmin) was similar to that of 3 or 15 arcmin]. Moreover, examining decoding performance as a function of the disparity difference between the stimuli (i.e., step size) suggested a further difference (**Figure 5***f*). In the majority of cortical areas, classification accuracy increased when the presented stimuli had a larger difference in depth (i.e., as a function of the disparity step size). However, classification accuracy in LO was unaffected by the disparity difference—that is, performance was similar whether the disparity planes were located close to the fixation point (±3 arcmin) or far away (±15 arcmin). This suggests that disparity representations in LO may be more categorical in reflecting local configurations (i.e., surfaces in front of or behind their neighbors) than the metric, or tuned, representations in dorsal areas. This result is clearly bounded by the experiential paradigm and analysis procedure (i.e., very small disparities would cause a drop off in performance in LO, and accuracies would saturate for very large disparities in other areas); nevertheless, subsequent single-unit recordings have supported the suggestion of metric versus categorical representation in dorsal versus ventral areas (Srivastava et al. 2009).

The suggestion of different forms of representation in the dorsal and ventral pathways is related to the study of binocular disparity expressed in different coordinate frames. Specifically, we can think about the disparity of, for example, two objects in a scene either (*a*) in a frame of reference anchored to the retina (absolute disparity) or (*b*) in terms of the difference in disparity between the two objects (relative disparity). When the eyes are moved, the objects change position on the retina, thereby changing the absolute disparity; however, the relative disparity between the objects will remain more or less constant. Humans are considerably more sensitive to relative depth (Westheimer 1979), which might suggest that the brain employs an explicit relative disparity reference frame. Single-unit recordings in V1 indicate absolute encoding (Cumming & Parker 1999), but V2 (Thomas et al. 2002) and V4 (Umeda et al. 2007) are intermediate between absolute

and relative frames of reference. Neri and colleagues (2004) used fMRI to test for representations of absolute versus relative disparity in human visual pathways. They found that absolute disparity changes drove responses in the dorsal pathway, and ventral areas (V4 and V8) were equally affected by changes in absolute or relative disparity information. These ventral responses may reflect a partial encoding of relative disparity signals or more abstracted template representations of local disparity configurations.

A consistent observation from human imaging has been the strong modulation of dorsal visual areas by disparity. Backus and colleagues (2001) first demonstrated a role for area V3A in human stereopsis, with subsequent work pointing to V3A, V3B/KO, V7, and the posterior parietal cortex (PPC) (Cottereau et al. 2011, 2012; Minini et al. 2010; Naganuma et al. 2005; Preston et al. 2008, 2009; Tyler et al. 2006). There is some evidence for lateralization to the right hemisphere (Durand et al. 2009, Georgieva et al. 2009, Ip et al. 2014, Nishida et al. 2001, Tsao et al. 2003), compatible with psychophysics (Durnford & Kimura 1971) and neuropsychology (Benton & Hécaen 1970, Carmon & Bechtoldt 1969, Hamsher 1978).

To probe disparity responses in the dorsal visual cortex, Goncalves and colleagues (2015) measured fMRI responses to parametric disparity manipulations (**Figure 6a**). By working at high field (7T) strengths, they closely sampled activity to test for spatial clustering in disparity representations. Work in macaques had suggested that disparity is systematically organized across the cortical surface (Chen et al. 2008, DeAngelis & Newsome 1999). Goncalves and colleagues (2015) found that disparity preferences (i.e., the disparity evoking the maximal response of a voxel) were spatially clustered and, importantly, were reproducible across imaging sessions (**Figure 6b**). They found clustering was particularly marked in dorsal areas (V3A, V3B/KO) (**Figure 6c**), with a likely spatial period of at least 3 mm across the cortical surface. This is compatible with estimates from the macaque brain (DeAngelis & Newsome 1999) scaled up to account for overall brain size.

Goncalves et al. (2015) sought to draw comparisons with electrophysiological measurements by fitting fMRI responses with Gabor models to find the voxels' peak response (referred to as the disparity preference) and response width. In early areas (V1, V2, V3d), the width of voxels' response profiles was not related to the overall disparity magnitude. However, in areas V3A and V3B/KO there was a significant relationship between the peak disparity and response width. This was intriguing because increases in the tuning width for larger disparity magnitudes are thought to be a characteristic of neural populations that underlie human stereoscopic judgments (Lehky & Sejnowski 1990, Stevenson et al. 1992). Goncalves et al. compared estimates of the relationship between peak disparity and tuning width from their fMRI data with those from psychophysical studies; this suggested a striking similarity between fMRI and psychophysical estimates in areas V3A and V3B/KO (**Figure 6d**).

How Are Local Signals Integrated into Global Surfaces?

Natural environments produce variations in disparity across space, for instance, when facing an inclined street (surface slant) or viewing the undulations of the human body (curvature). How does the brain translate from local estimates of disparity to more extensive surface properties?

The work so far described has considered rather simple disparity structures, such as step edges (Backus et al. 2001, Goncalves et al. 2015, Neri et al. 2004, Preston et al. 2008, Tsao et al. 2003, Tyler et al. 2006) or complex, disparity-defined curved shapes (Chandrasekaran et al. 2007, Georgieva et al. 2009, Minini et al. 2010, Nishida et al. 2001, Preston et al. 2008). For the purpose of understanding how local features are integrated into global surfaces, the former are too simple (i.e., local and global surface properties are perfectly correlated), but the latter are too complex (i.e., the parameters describing surface relief become too numerous), making it difficult to interpret changes in fMRI responses. Therefore, Ban & Welchman (2015) targeted slant (**Figure 7a**), an
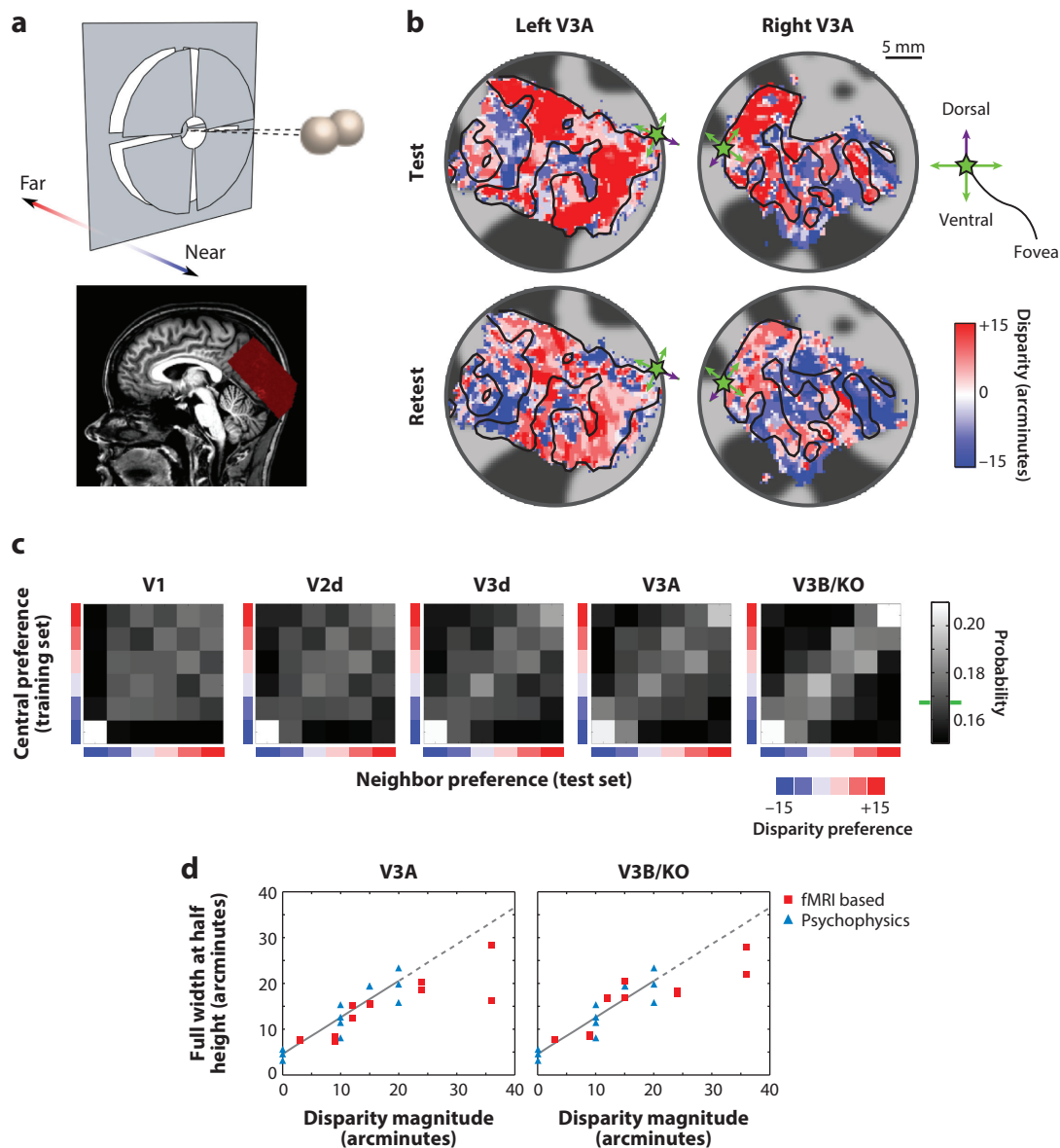
Figure 6

Results from Goncalves et al. (2015) using 7T functional magnetic resonance imaging (fMRI). (*a*) Parametric manipulations of disparity were used while collecting data from a small slab of voxels over the dorsal visual cortex. (*b*) Representation of the peak responses of individual voxels mapped onto the cortical surface of a single participant's V3A. Maps are shown for test and retest sessions. (*c*) Evidence for spatial clustering in fMRI responses emerges (*diagonal structure*) in higher dorsal areas. (*d*) Relation between peak response of a voxel and the width of the response profile, accompanied by historic estimates of this relationship from psychophysics (Stevenson et al. 1992). Figure adapted from Goncalves et al. (2015).
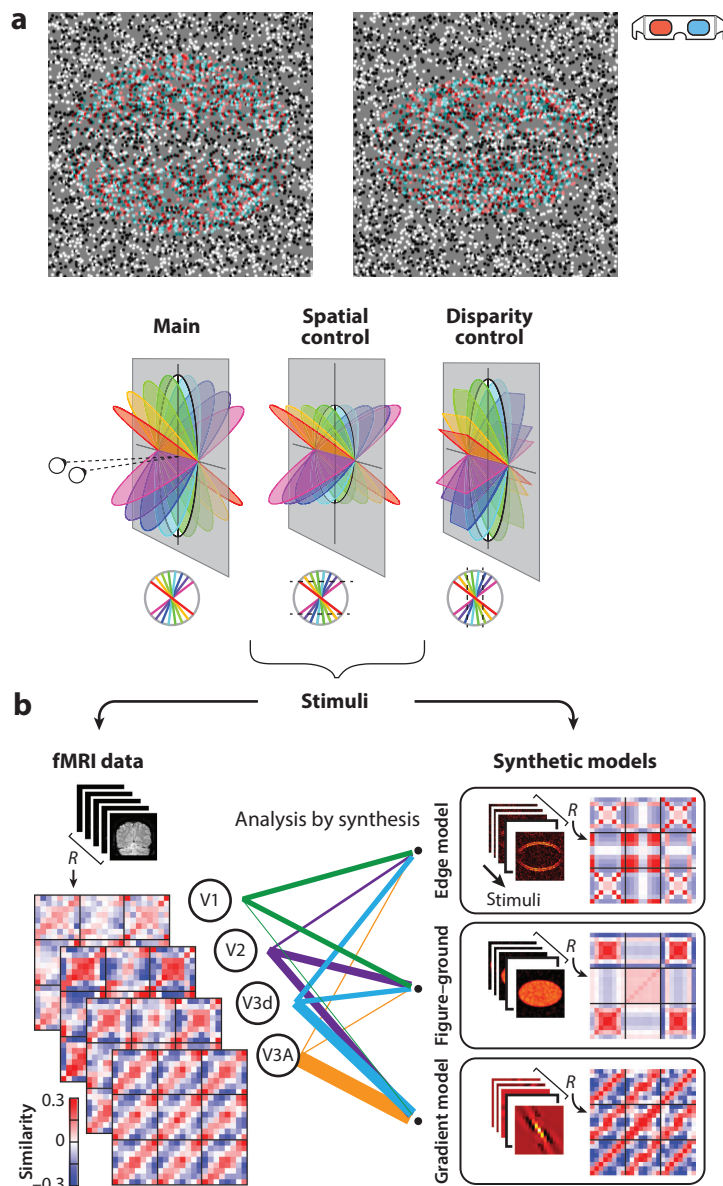
Figure 7

Decoding representations of slant. (*a*) Random dot stereograms for viewing using red-cyan anaglyph glasses depict slanted surfaces. As slant is changed, the projection size and range of disparities change. Therefore, functional magnetic resonance imaging (fMRI) responses were measured while controlling for changes in spatial projection and disparity range. (*b*) Overview of the analysis by synthesis approach. fMRI data were quantified using a similarity approach. An analogous approach was used for computer model simulations that implemented different processing operations. Data were analyzed by using model outputs as regressors for the empirical fMRI responses. Weights for the different regressors are shown schematically using line thickness. Figure adapted from Ban & Welchman (2015).

intermediate property that relates to the rotation of a plane away from frontoparallel (e.g., the steepness of a hillside path when walking). (Slant should not be confused with tilt, which is the rotation of a surface in the image plane, similar to the changing orientation of a clock's hands over time).

To analyze their data, Ban & Welchman (2015) developed an approach—analysis by synthesis— that sought to explain activity within a cortical area in terms of plausible computational processes. They combined voxel similarity analysis (Kriegeskorte et al. 2008) with synthetic models of elementary visual processes (i.e., edge detection, figure–ground segmentation, disparity-gradient detectors) to trace the development of representations across the processing hierarchy. They showed transformations in stimulus representations within the dorsal hierarchy from V1 to V3A (**Figure 7b**). In particular, responses in V2 had a strong component related to figure–ground segmentation, and V3A was strongly related to disparity gradients. Moreover, responses in V3A had characteristics similar to psychophysical judgments of slant. These results link to single-unit responses to slant in PPC (Rosenberg & Angelaki 2014, Rosenberg et al. 2013) that showed sensitivity to slant defined by both texture and disparity cues. V3A may extract disparity gradients that are then integrated with other depth cues at a further stage of processing [compatible with observations on surface tilt in the parietal (Tsutsui et al. 2002) and IT cortex (Liu et al. 2004)].

A test often used in single-unit studies has been to measure neural responses to slants and curvature while manipulating the overall disparity of the stimulus (i.e., changing its position in depth) (Janssen et al. 1999, 2000; Nguyenkim & DeAngelis 2003). This seeks to decouple the specific disparities defining the stimuli from descriptors that reflect the physical form of a viewed object (i.e., the objective 3D shape of the surface) and can be thought of as an extension to the concept of relative disparity encoding. Ban & Welchman (2015) manipulated the overall disparity of their stimuli and found that responses in V3A showed a degree of tolerance to changes in depth position. This suggested an intermediate form of surface representation that is beyond disparity gradients but not completely tolerant to changes in depth position. This is, perhaps, more compatible with single-unit results than it may first appear. In particular, neuronal spike rates are often strongly modulated by changes in the overall disparity of the stimuli (Janssen et al. 2000, Nguyenkim & DeAngelis 2003, Srivastava et al. 2009), indicating that although preferences for particular stimuli are maintained, neuronal responses are not invariant to the disparity structure defining the stimuli.

Ban & Welchman's (2015) study provided a first step in uncovering the human brain mechanisms that translate from isolated point estimates to descriptors of surface structure. However, they employed rather simple computational models, meaning that more (*a*) biologically realistic or (*b*) computationally sophisticated simulations may better capture the processing architectures within the human brain. Moreover, information processing hierarchies are not only spatial: The relative timing of responses in different areas (Cottereau et al. 2014, Srivastava et al. 2009) is likely to be critical in understanding the processing that underlies decisions about perceived depth structure.

How Do Disparity Representations Relate to Task Performance?

Observing strong responses to disparity is a useful starting point in uncovering the brain's depth processing architecture; however, to understand the purpose of such representations, it is useful to link responses to their functional use. Traditionally, we have learned about the importance of particular brain regions by studying the behavioral consequences of brain injury. For instance, neuropsychological testing indicated the importance of the right PPC for stereoscopic judgments (Benton & Hécaen 1970, Carmon & Bechtoldt 1969, Hamsher 1978) (although the lesions were not well characterized). More recent evidence has provided well-quantified descriptions of brain damage in conjunction with well-controlled psychophysical testing. Specifically, Read and

colleagues (2010) assessed patient DF, a visual form agnostic with extensive bilateral damage to lateral occipital cortex. They found that DF had remarkably good stereoscopic vision in spite of her injury, with deficits limited to performance on a relative disparity task. Moreover, this deficit was ameliorated after DF gained experience from testing. This suggests that DF's stereoscopic abilities depended on activity outside the LO cortex (plausibly within the dorsal stream), and that training enabled her to read out spared disparity representations in support of task performance.

Electrophysiological studies have established links between the activity of disparity sensitive neurons and perceptual judgments of depth in macaque area MT/V5 (Bradley et al. 1998, DeAngelis et al. 1998, Dodd et al. 2001). Subsequent work dissociated the neural substrates of two different disparity tasks in which performance was measured either by changing the signal-to-noise ratio (**Figure 8a**) or by titrating small differences between the depth of a target and its local neighborhood (**Figure 8a**). Before reviewing these findings, it is worth noting that these tasks typically involve differences in both disparity magnitude and in the stimulus manipulation used to affect task performance. This adds an interpretative complication, as differences in neural substrates may relate to (*a*) the task or (*b*) the disparity magnitude. These tasks have been widely discussed as "coarse" and "fine" tasks. However, these terms have specific meanings in relation to the magnitude of disparities related to binocular fusion (Norcia et al. 1985, Ogle 1952): Fine disparities are <30 arcmin. Therefore, I discuss them as, respectively, signal-in-noise tasks and feature-difference tasks.

Single-unit recordings have suggested that these tasks have different neural substrates: Area MT/V5 appears important for the signal-in-noise task, but V4 is implicated in the feature-difference task. In particular, for the signal-in-noise task, there are similarities between neural and psychophysical sensitivity in macaque MT/V5 (Uka & DeAngelis 2003, 2004), with electrical stimulation biasing psychophysical judgments (DeAngelis et al. 1998), and with inactivation reducing perceptual performance (Chowdhury & DeAngelis 2008). By contrast, electrical stimulation of MT/V5 had no measurable effect on a feature-difference task (Uka & DeAngelis 2006). In V4, performance on a feature-difference task was affected by microstimulation (Shiozaki et al. 2012), and neurons in both V4 and later ventral regions (IT) showed activity during the feature-difference task that predicted the perceptual choice (Shiozaki et al. 2012, Uka et al. 2005).

Based on these findings, Patten & Welchman (2015) recorded fMRI responses while human participants performed one or another of the tasks. They used similar magnitudes of disparity for the two tasks (within the classically defined fine range) and sampled activity across visual and parietal cortical areas. Their results pointed to similarities between performance on both tasks and fMRI responses in the PPC. This might suggest generalized disparity representations are manifest in higher portions of the dorsal stream or, alternatively, common areas are involved in reading out disparity signals to support task performance (Law & Gold 2008, Williams et al. 2003).

To provide a stronger test of the involvement of higher portions of the ventral and dorsal streams in performing these tasks, Chang and colleagues (2014) used transcranial magnetic stimulation (TMS) to disrupt neural activity within area LO versus PPC. They found a dissociation between TMS sites: Stimulation of PPC affected the signal-in-noise task in contrast to LO stimulation, which affected the feature-difference task (**Figure 8b**). The effects they measured corresponded to a worsening of the participants' disparity discrimination thresholds. Thus, TMS perturbed task performance but did not devastate judgments in the same way that lesions or invasive deactivations do (Chowdhury & DeAngelis 2008).

There is, however, a complication to the simple suggestion of dissociable neural loci between tasks. In particular, Chowdhury & DeAngelis (2008) showed in monkeys that the effects of MT/V5 inactivation on signal-in-noise tasks went away following a period of training on a
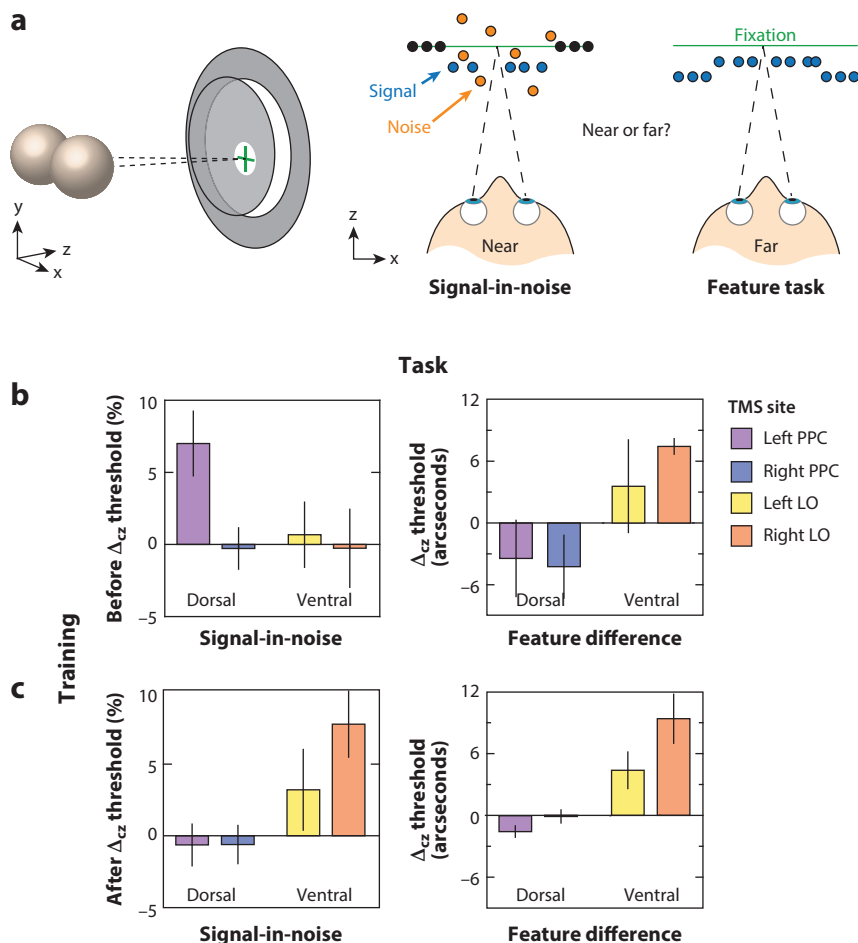
Figure 8

(*a*) Signal-in-noise task versus feature-difference task. For both tasks, the subject decides whether the central target is in front of or behind the surround. Performance is affected by changing the proportion of dots with randomly assigned disparities (i.e., signal-to-noise ratio) or titrating the disparity difference between the center and the surround. (*b,c*) Effects of dorsal versus ventral rTMS on signal-in-noise versus feature-difference task before (*b*) versus after (*c*) training. Results from Chang et al. (2014) show Δ_{cz}: the change in the psychophysical threshold under TMS relative to control site (cz) stimulation. Abbreviations: LO, lateral occipital; PPC, posterior parietal cortex; rTMS, repetitive transcranial magnetic stimulation.

feature-difference task. This suggested that perceptual experience changed the monkey's reliance on neural activity within area MT/V5 and that task performance was instead critically dependent on neural responses in another (unknown) location. This training effect drew strong parallels with psychophysical evidence that learning to discriminate small differences in the orientation of a target can boost participants' ability to detect a target embedded in noise (Dosher & Lu 1998, 2005; Dosher et al. 2013).

To test for the neuronal circuitry that might be engaged in the signal-in-noise task after training, Chang et al. (2014) included a perceptual learning dimension to their study. Following the initial TMS assessments, participants were trained on a feature-difference task for 3 days (which

produces benefits for the signal-in-noise task; see Chang et al. 2013). Thereafter, participants were reassessed under TMS. Chang et al. (2014) found that performance on the signal-in-noise task was no longer affected by TMS to PPC. Instead, stimulation of LO not only disrupted performance on the feature-difference task (as it had done before training) but also on the signal-in-noise task (**Figure 8***c*).

Chang et al. (2014) suggested that the same basic functional circuit was involved in the performance of both the signal-in-noise and feature-difference tasks. However, the performance of each task is limited by different processes: the ability to extract signals versus reading out stored feature representations. They suggested that training on the feature-difference task optimized the representations of disparity features in area LO. These boosted features facilitated figure–ground segmentation and the identification of targets in noise, diminishing the need for noise filtering by PPC. Thus, posttraining performance on the signal-in-noise task was no longer limited by signal extraction but rather by feature representations stored in the ventral cortex.

This interpretation places an emphasis on a network of cortical areas underlying task performance rather than on a single neural locus underlying perception. Nevertheless, it is unlikely that it provides a complete account. Evidence from patient DF (Read et al. 2010) suggested that she could learn to perform stereoscopic judgments well with a compromised ventral cortex. This might reflect a special case; however, these findings pose something of a general challenge to experimentalists used to working with well-trained human or animal observers whose visual systems have adapted away from their wild-type state.

THE NEURAL BASIS OF DEPTH CUE FUSION

Disparity is often regarded as the gold standard depth cue and has been a main focus when testing the neural mechanisms of 3D vision. Yet we seldom use disparity in isolation, and may ignore it all together (Bülthoff et al. 1998, Glennerster et al. 2006, Muryy et al. 2013). How does the brain integrate depth information obtained from different cues?

Let us start by considering some scenarios for representations within different parts of the brain. First, we might imagine an area that responds to only one type of cue (e.g., disparity): Changing this cue would result in different fMRI responses, but manipulating other depth information (e.g., texture, shading, or motion) would have no effect. As discussed at the outset, such a pure module seems unlikely. Second, we might imagine areas that contain subpopulations of neurons for different types of depth information (independence scenario). Changing one cue would change the fMRI response, and if we changed cues concurrently, we would anticipate bigger changes in the fMRI response. Finally, we might locate an area that, similar to perceptual judgments, integrates information from different cues into a unified depth representation (fusion scenario). In this case, changing the information provided by each cue would lead to fMRI changes if there were a change in the fused estimate (Welchman et al. 2005).

How could we identify and separate such representations? Before reviewing specific findings, it is worth revisiting our thoughts about the logic of brain imaging measurements.

Pitfalls in Using Imaging to Uncover Integration

A frequent starting point in designing an experiment to test for integration is to conceive of studying two cues (e.g., depth from disparity, depth from shading) and their relative controls (e.g., scrambled disparity, scrambled shading). We could make contrasts (disparity > scrambled disparity, shading > scrambled shading) to produce two maps that reveal putative brain locations favoring the impression of depth from each cue (**Figure 4***b*). To determine the site of integration,

we could overlay the two activity maps and reveal their intersection (**Figure 4b**). The logic of this approach is that the intersection reveals locations of convergence between the two depth cues.

However, there are limitations to this thinking. First, the area of intersection clearly depends on the statistical threshold of the individual contrasts, which, in turn, relates to the adequacy of the control stimulus (**Figure 4c**). In particular, using a poor contrasting null stimulus for one of the cues could lead to over- or underestimating the response related to the impression of depth. Second, because fMRI responses aggregate activity over many thousands of neurons, an overlap in fMRI responses for different cues could be caused by independent, but colocated, neural populations (**Figure 4c**).

An alternative approach might be to think about the strength of activity evoked when depth is defined by two cues relative to those evoked by either cue alone. Classic work on multisensory integration in the superior colliculus indicated that responses to bimodal stimulation can be superadditive—that is, they can exceed the response of adding each cue presented alone (Meredith & Stein 1983). So an alternative approach is to test for superadditive fMRI responses. However, for our purposes, there are a number of problems with this approach.

First, within the visual system it is difficult to talk about presenting cues alone or in isolation. By doing so, other cues are unwittingly placed into conflict with the cue under study. For illustration, consider an RDS (**Figure 5a**). The phenomenology of seeing depth from random black and white dots suggests a compelling means of studying depth perception using a pure disparity stimulus. However, RDSs often create large cue conflicts: The dot patterns specify that the viewed surface is flat (i.e., there is no texture gradient or element-size cue) in contrast to the disparity information. [A classic example of this is Erkelens & Collewijn's (1985) study that has been taken to indicate that the visual system is blind to changes in absolute disparity, but in fact it shows that disparity signals are overridden by the conflict with texture (Lugtigheid et al. 2011, Welchman et al. 2009)]. Similarly, isolating the shading cue often involves participants viewing stimuli binocularly, with the effect that the brain receives good evidence that the viewed surface is flat. Thus, it is difficult to talk about adding or subtracting depth cues because, despite our best intentions, the visual system will combine the available information, which can make us less sensitive (Hillis et al. 2002).

Second, the extent of additivity depends on the relationship between stimulus intensity and the neural response (Stanford et al. 2005)—that is, the neural transducer function (**Figure 4d**). Superadditive responses are typically observed around a single unit's response threshold, but more intense stimuli are additive or subadditive. Paradigms for fMRI typically involve presenting intense stimuli quite far from an observer's detection threshold (to ensure a measurable fMRI response), making it likely that responses would be outside the superadditive range. Potentially even more important, however, is that the relationship between neural activity and the BOLD fMRI response is complex (Heeger & Ress 2002) and imposes its own transducer function, which is unlikely to be a facsimile of the neural transducer function. Further, there may be unknown different transducer functions for different cues (**Figure 4d**), making predictions about the extent of additivity at the fMRI response level extremely difficult (see also James & Stevenson 2012, Laurienti et al. 2005).

Thus, fMRI responses may be superadditive because two cues are integrated to enhance neural responses (i.e., integration detected). But the same result (i.e., false positive) could be due to differences in the transducer function between cues; an increase in additional load due to monitoring two cues (Otto & Mamassian 2012); or there being multiple interactions, but not fusion, between neurons sensitive to the two depth cues. By contrast, responses may be subadditive because cues are integrated but the fMRI response has saturated; the sampled population of integrated neurons is small and, therefore, does not evoke a strong fMRI response; or the null stimulus is not well defined, weakening detection power (i.e., misses).

Testing for Cue Fusion

To provide alternative tests of cue fusion, Ban and colleagues (2012) developed a paradigm based on measuring the discriminability of fMRI responses from stimuli defined by disparity and motion cues. The basic intuition behind their approach is that it should be easier to tell two stimuli apart when differences between them are defined by two cues simultaneously. However, in contrast to the approach sketched above, they ensured that disparity and motion signals were always presented in the displays. In particular, they presented dot displays depicting a central target in front of or behind its surround (**Figure 9a**), in which the impression of depth was evoked by disparity or differences in the target's speed relative to its surround (relative motion), or both. Using this disparity–motion stimulus space, they created four conditions in which the target's near versus far position was defined by (*a*) disparity, D, in which motion indicated zero depth (i.e., flat); (*b*) relative motion, M, in which disparity indicated a flat surface; (*c*) both cues conveying consistent depths, DM (e.g., disparity = near; motion = near); or (*d*) incongruent cues, IC, in which cues indicated opposite depths (e.g., disparity = near; motion = far).

These stimuli were designed to distinguish between the fusion and independence mechanisms. However, a few steps of reasoning are required to understand precisely how we can tell them apart. First, we need to note that under both scenarios we expect the discriminability of near versus far configurations to be greater when both cues consistently indicate depth (DM condition). In the fusion case, this is because the variance of the estimators is lower: In **Figure 9b**, the fused distributions are more sharply peaked than the individual cue representations of the stimuli. In the independence case, this is because the estimators have a greater separation (i.e., the differently colored blobs in **Figure 9a** are furthest apart when projected orthogonally to the negative diagonal, as in **Figure 9b**). The improvement in the independence case corresponds to the quadratic sum of the separations along the disparity and motion axes [$DM = \sqrt{(D^2 + M^2)}$], which has an intuitive geometrical interpretation in terms of Pythagoras' theorem.

Second, we need to return to the idea that it is difficult to isolate individual cues. In particular, for "single" cues (i.e., D and M conditions), the fusion mechanism is compromised because differences between near versus far in one cue (e.g., disparity) are averaged together with the other (e.g., motion) that indicates a flat surface. This has the effect of depressing the sensitivity of the detection mechanism (**Figure 9c**). Having appreciated this point, we can use single cues to our advantage. In particular, the independence mechanism is not compromised by single cues: The lack of a difference along one axis can effectively be ignored (for a geometrical intuition of why this is so, consider that the hypotenuse can never be shorter than the catheti).

Figure 9

The depth of a central target can be defined in a space of disparity and motion. (*a*) Individual stimuli are represented as bivariate Gaussian distributions (*purple* versus *green blobs* for near versus far stimuli, respectively). A detector could sense depth along only one dimension (e.g., a disparity detector): Distinguishing the stimuli in this case depends on making a judgment using the marginal distribution (illustrated along the bottom and right-hand sides). (*b*) A fusion mechanism combines disparity and motion distributions into a single dimension; this reduces the variance of the combined estimate (*solid distributions*) relative to the components (*dotted distributions*). The independence mechanism finds the optimal separating boundary between the stimuli. This increases the separation between the distributions to improve discrimination performance. This corresponds to the quadratic sum of performance along the component axes (by the Pythagorean theorem this means greater separation along the diagonal). (*c*) Performance of the fusion versus independence mechanisms for the single-cue and incongruent-cue conditions. In both situations fusion performs worse than independence. (*d*) Decoding predictions for an area that responds on the basis of fusion versus independence. For fusion, conflict in single-cue disparity or motion conditions depresses sensitivity; this means that the empirical performance in the congruent disparity plus motion condition surpasses the predictions from the single-cue conditions. Abbreviations: D, disparity; M, relative motion. Figure adapted from Ban et al. (2012) and Dekker et al. (2015).
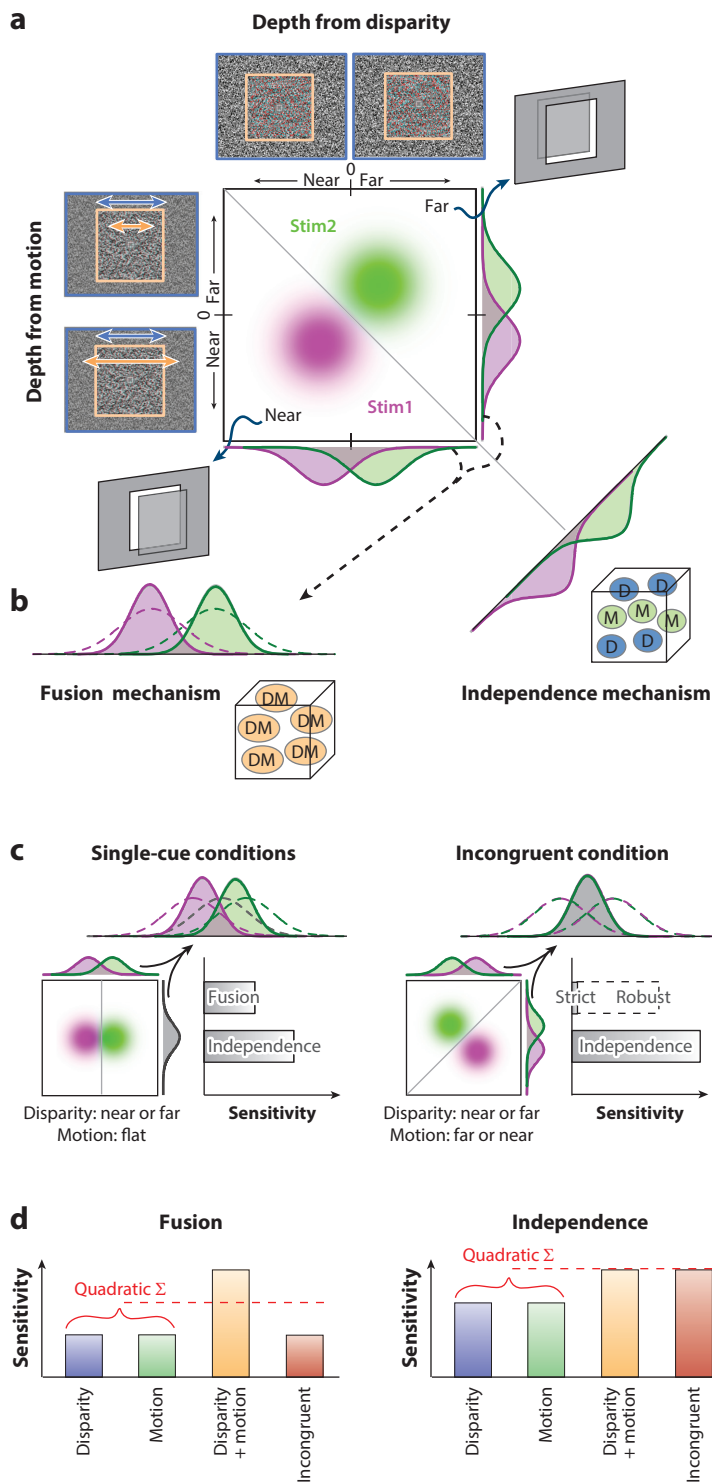
These differences in the performance of the fusion and independence mechanisms for single-cue stimuli allow us to separate them. In particular, we can use the discriminability of these conditions to make a prediction for performance in the consistent-cue (DM) condition. If an area represents independent depth estimates, discriminability in the DM condition should match the quadratic sum of the D and M conditions (**Figure 9d**). However, under fusion, because the quadratic sum prediction uses compromised D and M performance, the empirical performance in the DM case (where conflict is removed) will surpass the prediction (**Figure 9d**).

As a further test of fusion, we can assess performance in the case in which stimuli differ in both disparity and motion, but the cues specify the opposite depth configurations (the incongruent case). An independence mechanism should be unaffected by incongruency (**Figure 9c,d**), as quadratic summation ignores the sign of differences. However, a fusion mechanism would be affected: A strict fusion mechanism would be insensitive, whereas a robust mechanism (Landy et al. 1995) would revert to a single component (**Figure 9c,d**).

Ban and colleagues (2012) first measured performance psychophysically using these stimuli, finding that participants' performance was above the quadratic summation prediction for the DM condition and reverted to single-cue performance in the IC condition. This indicates that human judgments depend on robust cue fusion, consistent with earlier reports (Hillis et al. 2002, Knill & Saunders 2003). Ban and colleagues (2012) then used fMRI responses sampled from across the visual cortex to test for fusion at the neural level. Their results pointed to fusion in area V3B/KO (**Figure 10a**). In particular, in this area (but not elsewhere) DM decoding accuracies were reliably above the quadratic summation prediction, and decoding performance dropped to single-cue levels in the IC case.

As a further test, Ban et al. asked whether depth information provided by one cue (e.g., disparity) is diagnostic of depth indicated by the other (e.g., motion). They used a cross-cue transfer test whereby they trained their decoding algorithm to discriminate depth configurations using one cue and tested the classifier's predictions for data obtained when depth was indicated by the other. This result again suggested the importance of dorsal visual areas, and in particular V3B/KO (**Figure 10b**).

In a follow-up study, Dekker et al. (2015) tested for fMRI responses related to the integration of disparity and motion signals in children. Psychophysical tests had revealed that young children (<10.5 years) fail to fuse visual and multisensory signals (Gori et al. 2008; Nardini et al. 2008, 2010) as adults do, raising the question of whether visual responses related to cue fusion are also absent in younger children. Dekker et al. (2015) showed that children younger than 10.5 years did not show DM performance above quadratic summation nor did they show improved performance over ICs in V3B/KO. By contrast, children above this transition point showed adult-like responses. This suggests that the long developmental time course for cue fusion reflects changes occurring within the visual cortex and indicates a tight link between perceptual fusion and activity in V3B/KO.

This finding of a localized cortical region involved in integrating cues was compatible with an earlier suggestion that the KO region is specialized for the processing of depth structure (Tyler et al. 2006). Initially, this part of the visual cortex had been identified using contrasting stripes of motion, leading to its designation as the kinetic occipital region (Dupont et al. 1997, Van Oostende et al. 1997). However, based on tests with disparity-defined structures, Tyler et al. (2006) suggested (*a*) that it might be better thought of as an area that extracts generic depth representations and (*b*) that it is functionally separate from its neighbor V3B. Ban et al. (2012) found little to differentiate the processing within the retinotopically identified V3B and the functionally localized KO region, and treated this as one region of the cortex (activity from the KO localizer typically extends into V3B). An alternative classification scheme for this part of the cortex is the retinotopic localization of areas LO1 and LO2 (Larsson & Heeger 2006). **Figure 3** shows these labels overlaid with
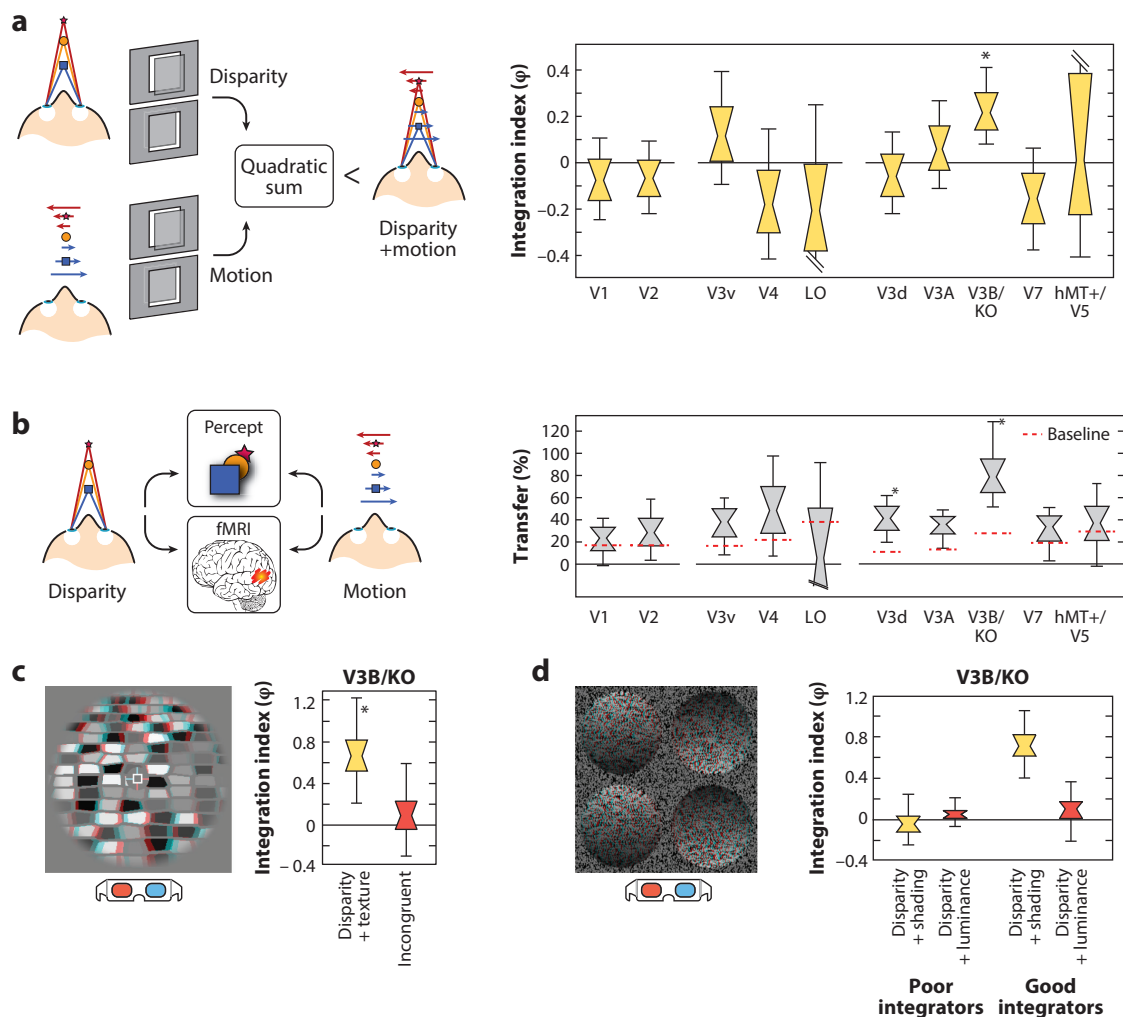
Figure 10

(*a*) Ban et al.'s (2012) results from the quadratic summation test across regions of interest. Results are shown as an integration index in which zero corresponds to performance in the congruent condition that matches the quadratic sum from single-cue conditions. Only in V3B/KO is performance reliably above the prediction. (*b*) Tests of transfer between depth cues. In V3B/KO, decoding performance across cues is very close to that within cues. (*c*) Stimuli and results from Murphy et al. (2013). V3B/KO shows evidence for fusion. (*d*) Stimuli and results from Dövencioğlu et al. (2013). V3B/KO was again implicated in fusion, with perceptual differences between participants reflected in differences in patterns of brain activity. Stimuli in panels *c* and *d* are designed for stereoscopic viewing using red-cyan anaglyph glasses. Abbreviations: fMRI, functional magnetic resonance imaging; KO, kinetic occipital.

V3B/KO, highlighting that there is overlap between LO1 and the V3B/KO designation. However, the retinotopic borders of LO1 and LO2 are often hard to identify, meaning that the precise delineation of these areas can be difficult (e.g., **Figure 3**), at least in the author's experience.

Testing Fusion for Other Cues with More Complex Shapes

Demonstrating that disparity and motion information are combined might plausibly reveal a special case of cue pairings. In particular, there are computational similarities between the cues

(disparity is frozen motion or motion-defrosted disparity, depending on your perspective), psychophysical results that suggest common stages of processing (Bradshaw & Rogers 1996, Domini et al. 2006, Nawrot & Blake 1989, Poom & Borjesson 1999), and imaging results indicate similarities between the processing of depth cues to motion and disparity (Murray et al. 2003, Peuskens et al. 2004, Vanduffel et al. 2002). Therefore, it was important to determine whether Ban et al.'s (2012) result was specific to disparity and motion or could be generalized to other cue pairings. Moreover, depth structure is often more complex than the simple depth order relationships used by Ban et al., so if V3B/KO provides a locus underlying behaviorally relevant fusion, it should also respond to more complex forms of surface structure.

To this end, Murphy et al. (2013) tested the combination of texture and disparity cues to surface slant. They found that V3B/KO produced fMRI responses that matched the expectations of cue fusion: When disparity and texture both signaled slant performance they exceeded the quadratic sum prediction, and this result was specific to congruent combinations of cues (**Figure 10c**).

Dövencioğlu et al. (2013) examined the combination of shading and disparity cues that depicted curved surfaces. They again found evidence for fusion in V3B/KO and that this activity was linked to psychophysical judgments (**Figure 10d**). In particular, they found that their participants varied considerably in their ability to integrate the shading and disparity signals: Some observers effectively ignored the shape information provided by shading when making their judgments. These behavioral differences between participants were reflected in the activity of V3B/KO, consistent with Dekker et al.'s (2015) observations.

Together, these findings suggest that a relatively circumscribed cortical area is involved in integrating cues to 3D structure. However, we should not interpret this as the site of a putative generic depth map. We simply do not know enough about the nature of the representations in this area, where they originate, which areas read them out, and how the different demands imposed by tasks such as controlling movements or recognizing objects might require different types of representations. Moreover, all of the tests have used disparity, so it is possible that activity in this area is peculiar to the use of disparity with other depth signals. Finally, it seems likely that V3B/KO also responds to other characteristics of viewed objects (Sun et al. 2014, Wada et al. 2014), so its role is unlikely to be limited simply to inferring 3D structure.

FUTURE DIRECTIONS

The development of imaging approaches during the past two decades has dramatically improved our understanding of depth representations in the human brain. Yet there is still a considerable gap between the detailed models we have for the early cortical processes that measure disparity (e.g., binocular energy model) and the higher cortical areas that represent fused depth structure for potentially multiple purposes. Further, we have a poor understanding of the computational processes that underlie the brain's extraction of depth from cues other than disparity or motion. The analysis-by-synthesis approach that seeks to expose the evolution of visual representations within a processing hierarchy seems promising in helping to move understanding from localization (i.e., where?) to computational processes (i.e., how?). Moreover, advances in machine learning provide an opportunity to move from handcrafted, intuitive models to more biologically grounded neural networks, whose structure has been shaped by exposure to natural images (Kriegeskorte 2015). Marrying better computational models to higher-resolution samples of brain activity is likely to be important in understanding the overall architecture of depth representations in the human brain.

A further challenge is to move beyond a static understanding of the functional architecture to one that incorporates information about the temporal dynamics of representations. Recent EEG

work has provided a glimpse of the relative timing of activity within the cortex that underlies the perceptual interpretation of depth (Cottereau et al. 2014). This approach is based on understanding network-wide dynamics that may support processes for enhancing and reading out depth representations, for instance, as suggested by the studies that point to plasticity in the circuitry that underlies task performance (Chang et al. 2014, Chowdhury & DeAngelis 2008). However, interactions are also likely to be critical at a local scale. Imaging techniques are starting to reveal processing architectures within the laminar structure of the cortex (Muckli et al. 2015), offering the potential to study how depth computations are shaped by combinations of feedforward, lateral, and feedback inputs.

CONCLUDING REMARKS

In this review, we have seen the potential role for a range of different cortical areas in computing the structure of the surrounding environment based on visual signals. The results reviewed suggest there may be more modularization for depth processing than we might initially have supposed. In particular, higher dorsal areas of the visual cortex appear highly responsive to depth, for both disparity and its combination with other signals. Moreover, there is evidence for differences in the nature of the representations between the higher dorsal areas and ventral area LO.

Previous discussions of the literature have suggested a distinction between the dorsal and ventral cortices in terms of a division of labor between absolute versus relative signals (Neri 2005, Parker 2007) or different magnitudes of disparity (Tyler 1990) and task manipulations (Roe et al. 2007). Here, I suggest a different way of capturing these differences that is centered on different computational demands.

Under our discussion of cue integration, we focused on two means for improving the sensitivity of an observer's judgments: increasing the separation between two estimates (i.e., the independence scheme in **Figure 9b**) versus reducing the variance of the estimators (the fusion scheme in **Figure 9b**). These two modes of operation may be exploited for different types of tasks. If a body movement is required, the brain is best served by fusing the available information to obtain an estimate of the scene that is unbiased and has low variance. Such a representation would be particular to the viewing situation (that is, highly specific) and variant under the manipulation of individual cues. This provides the best metric information about the scene that is specific to the current view. By contrast, recognition tasks are best served by maximizing the separation of objects in a high-dimensional feature space while ignoring uninformative dimensions. Such a mechanism would support invariant performance by discarding irrelevant, or nuisance, scene parameters (e.g., uninformative cues), yet it may be highly uncertain about the particular structure of the scene. Perceptual learning may have a key role in this process by using ventral circuits to store abstracted versions of configurations extracted by dorsal visual areas to support the recognition and the performance of learned tasks.

SUMMARY POINTS

1. 3D vision depends on an interference process that resolves locally ambiguous signals.

2. The brain exploits multiple different depth signals and uses 3D information to support a wide range of behaviors. These multiple routes in and routes out need to be considered when studying neural responses to 3D structure.

3. At a mechanistic level, depth processing is broadly understood within a modular architecture in which different cues are processed quasi-independently. At the neural level, depth signals are represented at multiple stages within the visually responsive cortex.

4. Brain imaging methods allow us to image neural responses to 3D structure. However, careful consideration needs to be given to the composition of the control stimuli and the difficulties of isolating individual cues.

5. In the cortex, fMRI responses to disparity are widespread, with strong responses found around area V3A. The nature of the disparity representations appears different in dorsal versus ventral areas, with highly specific and clustered representations in dorsal areas.

6. Extracting spatial variations of disparity is likely to involve a dorsal hierarchy, with fMRI evidence suggesting an important role for V3A in extracting disparity gradients. Responses in this area share characteristics with psychophysical judgments of depth.

7. Perceptual judgments of depth are likely to depend on activity in a network of areas that includes the ventral LO and PPC. Extensive training appears to refine responses and shift the limits on perceptual judgments from parietal to ventral areas.

8. Responses in area V3B/KO appear to be closely related to the fusion of different depth cues. A range of different cues appears to be integrated in this area, and developmental changes in the ability of children to fuse cues are mirrored by changes in activity in this area.

FUTURE ISSUES

1. What is the computational architecture for processing depth cues other than disparity? Are there processes analogous to the binocular energy model?

2. What are the hierarchical processes that converge in V3B/KO for cue fusion? How are responses in this region read out to support different tasks?

3. Can better artificial models improve our understanding of processing within the human brain? Integrating deep neural networks with fMRI (analysis by synthesis) may offer us new insights into hierarchical computations within the human brain.

4. What are the dynamics of processing between different regions of the cortex?

5. What is the impact of training on the responses measured in different cortical areas? To what extent are practiced observers representative of wild-type depth processing?

6. How will higher field strengths enable us to understand depth representations within the human brain? Can we uncover systematic maps that are consistent between different cues? What is the relationship between such maps and retinotopic organization?

7. What are the interactions within local areas of the cortex that support 3D perception? Can we uncover computations in different laminar layers?

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

Babington-Smith C. 1958. *Evidence in Camera*. London: Chatto & Windus

Backus BT, Fleet DJ, Parker AJ, Heeger DJ. 2001. Human cortical activity correlates with stereoscopic depth perception. *J. Neurophysiol.* 86(4):2054–68

Badcock DR, Schor CM. 1985. Depth-increment detection function for individual spatial channels. *J. Opt. Soc. Am. A* 2(7):1211–16

Ban H, Preston TJ, Meeson A, Welchman AE. 2012. The integration of motion and disparity cues to depth in dorsal visual cortex. *Nat. Neurosci.* 15(4):636–43

Ban H, Welchman AE. 2015. fMRI analysis-by-synthesis reveals a dorsal hierarchy that extracts surface slant. *J. Neurosci.* 35(27):9823–35

Benton AL, Hécaen H. 1970. Stereoscopic vision in patients with unilateral cerebral disease. *Neurology* 20(11):1084

Blake A, Bülthoff H. 1990. Does the brain know the physics of specular reflection? *Nature* 343(6254):165–68

Bradley DC, Chang GC, Andersen RA. 1998. Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature* 392(6677):714–17

Bradshaw MF, Rogers BJ. 1996. The interaction of binocular disparity and motion parallax in the computation of depth. *Vis. Res.* 36(21):3457–68

Bridge H, Parker AJ. 2007. Topographical representation of binocular depth in the human visual cortex using fMRI. *J. Vis.* 7(14):15

Bülthoff HH, Mallot HA. 1988. Integration of depth modules: stereo and shading. *J. Opt. Soc. Am. A* 5(10):1749–58

Bülthoff I, Bülthoff H, Sinha P. 1998. Top-down influences on stereoscopic depth-perception. *Nat. Neurosci.* 1(3):254–57

Burge J, Geisler WS. 2014. Optimal disparity estimation in natural stereo images. *J. Vis.* 14(2):1

Carmon A, Bechtoldt HP. 1969. Dominance of the right cerebral hemisphere for stereopsis. *Neuropsychologia* 7(1):29–39

Chandrasekaran C, Canon V, Dahmen JC, Kourtzi Z, Welchman AE. 2007. Neural correlates of disparity-defined shape discrimination in the human brain. *J. Neurophysiol.* 97(2):1553–65

Chang DHF, Kourtzi Z, Welchman AE. 2013. Mechanisms for extracting a signal from noise as revealed through the specificity and generality of task training. *J. Neurosci.* 33(27):10962–71

Chang DHF, Mevorach C, Kourtzi Z, Welchman AE. 2014. Training transfers the limits on perception from parietal to ventral cortex. *Curr. Biol.* 24(20):2445–50

Chen G, Lu HD, Roe AW. 2008. A map for horizontal disparity in monkey V2. *Neuron* 58(3):442–50

Chowdhury SA, DeAngelis GC. 2008. Fine discrimination training alters the causal contribution of macaque area MT to depth perception. *Neuron* 60(2):367–77

Clark J, Yuille AL. 1990. *Data Fusion for Sensory Information Processing Systems*. Boston: Kluwer Acad.

Cogan AI, Lomakin AJ, Rossi AF. 1993. Depth in anticorrelated stereograms: effects of spatial density and interocular delay. *Vis. Res.* 33(14):1959–75

Cottereau BR, Ales JM, Norcia AM. 2014. The evolution of a disparity decision in human visual cortex. *NeuroImage* 92:193–206

Cottereau BR, McKee SP, Ales JM, Norcia AM. 2011. Disparity-tuned population responses from human visual cortex. *J. Neurosci.* 31(3):954–65

Cottereau BR, McKee SP, Ales JM, Norcia AM. 2012. Disparity-specific spatial interactions: evidence from EEG source imaging. *J. Neurosci.* 32(3):826–40

Cumming BG, DeAngelis GC. 2001. The physiology of stereopsis. *Annu. Rev. Neurosci.* 24:203–38

Cumming BG, Parker AJ. 1997. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* 389(6648):280–83

Cumming BG, Parker AJ. 1999. Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J. Neurosci.* 19(13):5602–18

DeAngelis G, Cumming B, Newsome W. 1998. Cortical area MT and the perception of stereoscopic depth. *Nature* 394(6694):677–80

DeAngelis GC, Newsome WT. 1999. Organization of disparity-selective neurons in macaque area MT. *J. Neurosci.* 19(4):1398–415

Dekker TM, Ban H, van der Velde B, Sereno MI, Welchman AE, Nardini M. 2015. Late development of cue integration is linked to sensory fusion in cortex. *Curr. Biol.* 25(21):2856–61

Dodd J, Krug K, Cumming B, Parker AJ. 2001. Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area. *J. Neurosci.* 21(13):4809–21

Domini F, Caudek C, Tassinari H. 2006. Stereo and motion information are not independently processed by the visual system. *Vis. Res.* 46(11):1707–23

Dosher BA, Jeter P, Liu J, Lu ZL. 2013. An integrated reweighting theory of perceptual learning. *PNAS* 110(33):13678–83

Dosher BA, Lu ZL. 1998. Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *PNAS* 95(23):13988–93

Dosher BA, Lu ZL. 2005. Perceptual learning in clear displays optimizes perceptual expertise: learning the limiting process. *PNAS* 102(14):5286–90

Dosher BA, Sperling G, Wurst SA. 1986. Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vis. Res.* 26(6):973–90

Dövencioğlu D, Ban H, Schofield AJ, Welchman AE. 2013. Perceptual integration for qualitatively different 3-D cues in the human brain. *J. Cogn. Neurosci.* 25(9):1527–41

Dupont P, De Bruyn B, Vandenberghe R, Rosier AM, Michiels J, et al. 1997. The kinetic occipital region in human visual cortex. *Cereb. Cortex* 7(3):283–92

Durand J-B, Peeters R, Norman JF, Todd JT, Orban GA. 2009. Parietal regions processing visual 3D shape extracted from disparity. *NeuroImage* 46(4):1114–26

Durnford M, Kimura D. 1971. Right hemisphere specialization for depth perception reflected in visual field differences. *Nature* 231(5302):394–95

Erkelens C, Collewijn H. 1985. Eye-movements and stereopsis during dichoptic viewing of moving random-dot stereograms. *Vis. Res.* 25(11):1689–1700

Ernst MO, Banks MS. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415(6870):429–33

Fleet DJ, Wagner H, Heeger DJ. 1996. Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vis. Res.* 36(12):1839–57

Fleming RW, Holtmann-Rice D, Bülthoff HH. 2011. Estimation of 3D shape from image orientations. *PNAS* 108(51):20438–43

Fleming RW, Torralba A, Adelson EH. 2004. Specular reflections and the perception of shape. *J. Vis.* 4(9):798–820

Georgieva SS, Peeters R, Kolster H, Todd JT, Orban GA. 2009. The processing of three-dimensional shape from disparity in the human brain. *J. Neurosci.* 29(3):727–42

Georgieva SS, Todd JT, Peeters R, Orban GA. 2008. The extraction of 3D shape from texture and shading in the human brain. *Cereb. Cortex* 18(10):2416–38

Gerardin P, Kourtzi Z, Mamassian P. 2010. Prior knowledge of illumination for 3D perception in the human brain. *PNAS* 107(37):16309–14

Gillebert CR, Schaeverbeke J, Bastin C, Neyens V, Bruffaerts R, et al. 2015. 3D shape perception in posterior cortical atrophy: a visual neuroscience perspective. *J. Neurosci.* 35(37):12673–92

Glennerster A, Tcheang L, Gilson SJ, Fitzgibbon AW, Parker AJ. 2006. Humans ignore motion and stereo cues in favor of a fictional stable world. *Curr. Biol.* 16(4):428–32

Goncalves NR, Ban H, Sánchez-Panchuelo RM, Francis ST, Schluppeck D, Welchman AE. 2015. 7 Tesla fMRI reveals systematic functional organization for binocular disparity in dorsal visual cortex. *J. Neurosci.* 35(7):3056–72

Gonzalez F, Perez R. 1998. Neural mechanisms underlying stereoscopic vision. *Prog. Neurobiol.* 55(3):191–224

Gori M, Del Viva M, Sandini G, Burr DC. 2008. Young children do not integrate visual and haptic form information. *Curr. Biol.* 18(9):694–98

Haefner RM, Cumming BG. 2008. Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model. *Neuron* 57(1):147–58

Hamsher KD. 1978. Stereopsis and unilateral brain disease. *Investig. Ophthalmol. Vis. Sci.* 17(4):336–43

Heeger DJ, Ress D. 2002. What does fMRI tell us about neuronal activity? *Nat. Rev. Neurosci.* 3(2):142–51

Hibbard P. 2007. A statistical model of binocular disparity. *Vis. Cogn.* 15(2):149–65

Hillis JM, Ernst MO, Banks MS, Landy MS. 2002. Combining sensory information: mandatory fusion within, but not between, senses. *Science* 298(5598):1627–30

Hubel DH, Wiesel TN. 1970. Stereoscopic vision in macaque monkey: cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature* 225(5227):41–42

Ip IB, Minini L, Dow J, Parker AJ, Bridge H. 2014. Responses to interocular disparity correlation in the human cerebral cortex. *Ophthalmic Physiol. Opt.* 34(2):186–98

Jacobs RA. 1999. Optimal integration of texture and motion cues to depth. *Vis. Res.* 39(21):3621–29

James TW, Stevenson RA. 2012. The use of fMRI to assess multisensory integration. In *Frontiers in the Neural Bases of Multisensory Processes*, ed. MM Murray, MT Wallace, pp. 131–46. Boca Raton, FL: CRC

Janssen P, Vogels R, Liu Y, Orban GA. 2003. At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron* 37(4):693–701

Janssen P, Vogels R, Orban GA. 1999. Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *PNAS* 96(14):8217–22

Janssen P, Vogels R, Orban GA. 2000. Three-dimensional shape coding in inferior temporal cortex. *Neuron* 27(2):385–97

Johnston EB, Cumming BG, Landy MS. 1994. Integration of stereopsis and motion shape cues. *Vis. Res.* 34(17):2259–75

Johnston EB, Cumming BG, Parker AJ. 1993. Integration of depth modules: stereopsis and texture. *Vis. Res.* 33(5–6):813–26

Julesz B. 1971. *Foundations of Cyclopean Perception*. Chicago: Univ. Chicago Press

Knill D, Saunders J. 2003. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vis. Res.* 43(24):2539–58

Kriegeskorte N. 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* 1:417–46

Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2:4

Landy M, Maloney L, Johnston EB, Young M. 1995. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vis. Res.* 35(3):389–412

Larsson J, Heeger DJ. 2006. Two retinotopic visual areas in human lateral occipital cortex. *J. Neurosci.* 26(51):13128–42

Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE. 2005. On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp. Brain Res.* 166(3–4):289–97

Law C-T, Gold JI. 2008. Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. *Nat. Neurosci.* 11(4):505–13

Lehky SR, Sejnowski TJ. 1990. Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *J. Neurosci.* 10(7):2281–99

Liu Y, Vogels R, Orban GA. 2004. Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *J. Neurosci.* 24(15):3795–3800

Lugtigheid AJ, Brenner E, Welchman AE. 2011. Speed judgments of three-dimensional motion incorporate extraretinal information. *J. Vis.* 11(13):1

Marr D. 1976. Early processing of visual information. *Philos. Trans. R. Soc. Lond. Ser. B* 275(942):483–519

Marr D, Poggio T. 1976. Cooperative computation of stereo disparity. *Science* 194(4262):283–87

Marr D, Poggio T. 1979. A computational theory of human stereo vision. *Proc. R. Soc. Lond. Ser. B* 204(1156):301–28

Meredith MA, Stein BE. 1983. Interactions among converging sensory inputs in the superior colliculus. *Science* 221(4608):389–91

Minini L, Parker AJ, Bridge H. 2010. Neural modulation by binocular disparity greatest in human dorsal visual stream. *J. Neurophysiol.* 104(1):169–78

Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, et al. 2015. Contextual feedback to superficial layers of V1. *Curr. Biol.* 25(20):2690–95

Murphy AP, Ban H, Welchman AE. 2013. Integration of texture and disparity cues to surface slant in dorsal visual cortex. *J. Neurophysiol.* 110(1):190–203

Murray SO, Olshausen BA, Woods DL. 2003. Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cereb. Cortex* 13(5):508–16

Muryy AA, Welchman AE, Blake A, Fleming RW. 2013. Specular reflections and the estimation of shape from binocular disparity. *PNAS* 110(6):2413–18

Naganuma T, Nose I, Inoue K, Takemoto A, Katsuyama N, Taira M. 2005. Information processing of geometrical features of a surface based on binocular disparity cues: an fMRI study. *Neurosci. Res.* 51(2):147–55

Nardini M, Bedford R, Mareschal D. 2010. Fusion of visual cues is not mandatory in children. *PNAS* 107(39):17041–46

Nardini M, Jones P, Bedford R, Braddick O. 2008. Development of cue integration in human navigation. *Curr. Biol.* 18(9):689–93

Nawrot M, Blake R. 1989. Neural integration of information specifying structure from stereopsis and motion. *Science* 244(4905):716–18

Neri P. 2005. A stereoscopic look at visual cortex. *J. Neurophysiol.* 93(4):1823–26

Neri P, Bridge H, Heeger DJ. 2004. Stereoscopic processing of absolute and relative disparity in human visual cortex. *J. Neurophysiol.* 92(3):1880–91

Nguyenkim JD, DeAngelis GC. 2003. Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J. Neurosci.* 23(18):7117–28

Nishida Y, Hayashi O, Iwami T, Kimura M, Kani K, et al. 2001. Stereopsis-processing regions in the human parieto-occipital cortex. *NeuroReport* 12(10):2259–63

Norcia AM, Suiter EE, Tyler CW. 1985. Electrophysiological evidence for the existence of coarse and fine disparity mechanisms in human. *Vis. Res.* 25(11):1603–11

Ogle KN. 1932. An analytical treatment of the longitudinal horopter, its measurement and application to related phenomena, especially to the relative size and shape of the ocular images. *J. Opt. Soc. Am.* 22(12):665–728

Ogle KN. 1952. Disparity limits of stereopsis. *AMA Arch. Ophthalmol.* 48(1):50–60

Ohzawa I, DeAngelis GC, Freeman RD. 1990. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249(4972):1037–41

Orban GA. 2011. The extraction of 3D shape in the visual system of human and nonhuman primates. *Annu. Rev. Neurosci.* 34:361–88

Orban GA, Janssen P, Vogels R. 2006. Extracting 3D structure from disparity. *Trends Neurosci.* 29(8):466–73

Otto TU, Mamassian P. 2012. Noise and correlations in parallel perceptual decision making. *Curr. Biol.* 22(15):1391–96

Parker AJ. 2007. Binocular depth perception and the cerebral cortex. *Nat. Rev. Neurosci.* 8(5):379–91

Patten ML, Welchman AE. 2015. fMRI activity in posterior parietal cortex relates to the perceptual use of binocular disparity for both signal-in-noise and feature difference tasks. *PLOS ONE* 10(11):e0140696

Peuskens H, Claeys KG, Todd JT, Norman JF, Van Hecke P, Orban GA. 2004. Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. *J. Cogn. Neurosci.* 16(4):665–82

Pirenne MH. 1970. *Optics, Painting and Photography*. Cambridge, UK: Cambridge Univ. Press

Poggio GF, Gonzalez F, Krause F. 1988. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J. Neurosci.* 8(12):4531–50

Ponce CR, Lomber SG, Born RT. 2008. Integrating motion and depth via parallel pathways. *Nat. Neurosci.* 11(2):216–23

Poom L, Borjesson E. 1999. Perceptual depth synthesis in the visual system as revealed by selective adaptation. *J. Exp. Psychol. Hum. Percept. Perform.* 25(2):504–17

Preston TJ, Kourtzi Z, Welchman AE. 2009. Adaptive estimation of three-dimensional structure in the human brain. *J. Neurosci.* 29(6):1688–98

Preston TJ, Li S, Kourtzi Z, Welchman AE. 2008. Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. *J. Neurosci.* 28(44):11315–27

Read JCA, Cumming BG. 2007. Sensors for impossible stimuli may solve the stereo correspondence problem. *Nat. Neurosci.* 10(10):1322–28

Read JCA, Phillipson GP, Serrano-Pedraza I, Milner AD, Parker AJ. 2010. Stereoscopic vision in the absence of the lateral occipital cortex. *PLOS ONE* 5(9):e12608

Regan D, Spekreijse H. 1970. Electrophysiological correlate of binocular depth perception in man. *Nature* 225(5227):92–94

Roe AW, Parker AJ, Born RT, Deangelis GC. 2007. Disparity channels in early vision. *J. Neurosci.* 27(44):11820–31

Rosenberg A, Angelaki DE. 2014. Reliability-dependent contributions of visual orientation cues in parietal cortex. *PNAS* 111(50):18043–48

Rosenberg A, Cowan NJ, Angelaki DE. 2013. The visual representation of 3D object orientation in parietal cortex. *J. Neurosci.* 33(49):19352–61

Shiozaki HM, Tanabe S, Doi T, Fujita I. 2012. Neural activity in cortical area V4 underlies fine disparity discrimination. *J. Neurosci.* 32(11):3830–41

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

Stanford TR, Quessy S, Stein BE. 2005. Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J. Neurosci.* 25(28):6499–508

Stevenson SB, Cormack LK, Schor CM, Tyler CW. 1992. Disparity tuning in mechanisms of human stereopsis. *Vis. Res.* 32(9):1685–94

Sun H-C, Ban H, Di Luca M, Welchman AE. 2014. fMRI evidence for areas that process surface gloss in the human visual cortex. *Vis. Res.* 109:149–57

Tanabe S, Umeda K, Fujita I. 2004. Rejection of false matches for binocular correspondence in macaque visual cortical area V4. *J. Neurosci.* 24(37):8170–80

Thomas OM, Cumming BG, Parker AJ. 2002. A specialization for relative disparity in V2. *Nat. Neurosci.* 5(5):472–78

Ts'o DY, Roe AW, Gilbert CD. 2001. A hierarchy of the functional organization for color, form and disparity in primate visual area V2. *Vis. Res.* 41(10–11):1333–49

Tsao DY, Vanduffel W, Sasaki Y, Fize D, Knutsen TA, et al. 2003. Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron* 39(3):555–68

Tsutsui K-I, Sakata H, Naganuma T, Taira M. 2002. Neural correlates for perception of 3D surface orientation from texture gradient. *Science* 298(5592):409–12

Tyler CW. 1990. A stereoscopic view of visual processing streams. *Vis. Res.* 30(11):1877–95

Tyler CW, Likova LT, Kontsevich LL, Wade AR. 2006. The specificity of cortical region KO to depth structure. *NeuroImage* 30(1):228–38

Uka T, DeAngelis GC. 2003. Contribution of middle temporal area to coarse depth discrimination: comparison of neuronal and psychophysical sensitivity. *J. Neurosci.* 23(8):3515–30

Uka T, DeAngelis GC. 2004. Contribution of area MT to stereoscopic depth perception: Choice-related response modulations reflect task strategy. *Neuron* 42(2):297–310

Uka T, DeAngelis GC. 2006. Linking neural representation to function in stereoscopic depth perception: roles of the middle temporal area in coarse versus fine disparity discrimination. *J. Neurosci.* 26(25):6791–802

Uka T, Tanabe S, Watanabe M, Fujita I. 2005. Neural correlates of fine depth discrimination in monkey inferior temporal cortex. *J. Neurosci.* 25(46):10796–802

Umeda K, Tanabe S, Fujita I. 2007. Representation of stereoscopic depth based on relative disparity in macaque area V4. *J. Neurophysiol.* 98(1):241–52

Van Oostende S, Sunaert S, Van Hecke P, Marchal G, Orban GA. 1997. The kinetic occipital (KO) region in man: an fMRI study. *Cereb. Cortex* 7(7):690–701

Vanduffel W, Fize D, Peuskens H, Denys K, Sunaert S, et al. 2002. Extracting 3D from motion: differences in human and monkey intraparietal cortex. *Science* 298(5592):413–15

Wada A, Sakano Y, Ando H. 2014. Human cortical areas involved in perception of surface glossiness. *NeuroImage* 98:243–57

Wallach H. 1935. Über visuell wahrgenommene Bewegungsrichtung. *Psychol. Forsch.* 20:325–80

Wallach H, O'Connell DN. 1953. The kinetic depth effect. *J. Exp. Psychol.* 45(4):205–17

Watt SJ, Bradshaw MF. 2002. Binocular information in the control of prehensile movements in multiple-object scenes. *Spat. Vis.* 15(2):141–55

Welchman AE, Deubelius A, Conrad V, Bülthoff HH, Kourtzi Z. 2005. 3D shape perception from combined depth cues in human visual cortex. *Nat. Neurosci.* 8(6):820–27

Welchman AE, Harris JM, Brenner E. 2009. Extra-retinal signals support the estimation of 3D motion. *Vis. Res.* 49(7):782–89

Welchman AE, Kourtzi Z. 2013. Linking brain imaging signals to visual perception. *Vis. Neurosci.* 30(5–6):229–41

Westheimer G. 1979. Cooperative neural processes involved in stereoscopic acuity. *Exp. Brain Res.* 36(3):585–97

Williams ZM, Elfar JC, Eskandar EN, Toth LJ, Assad JA. 2003. Parietal activity and the perceived direction of ambiguous apparent motion. *Nat. Neurosci.* 6(6):616–23

ANNUAL REVIEWS
Connect With Our Experts
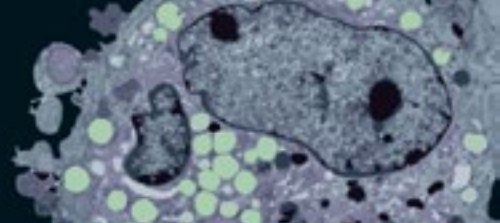
ONLINE NOW!

New From Annual Reviews:

Annual Review of Cancer Biology
cancerbio.annualreviews.org · Volume 1 · March 2017

Co-Editors: Tyler Jacks, *Massachusetts Institute of Technology*

Charles L. Sawyers, *Memorial Sloan Kettering Cancer Center*

The *Annual Review of Cancer Biology* reviews a range of subjects representing important and emerging areas in the field of cancer research. The *Annual Review of Cancer Biology* includes three broad themes: Cancer Cell Biology, Tumorigenesis and Cancer Progression, and Translational Cancer Science.

ANNUAL REVIEWS | CONNECT WITH OUR EXPERTS

650.493.4400/800.523.8635 (US/CAN)
www.annualreviews.org | service@annualreviews.org

$\overset{A}{R}$

Contents

Errata

An online log of corrections to *Annual Review of Vision Science* articles may be found at
http://www.annualreviews.org/errata/vision