ANNUAL REVIEWS

Annual Review of Vision Science

Neuronal Representations Supporting Three-Dimensional Vision in Nonhuman Primates

Ari Rosenberg,[1] Lowell W. Thompson,[1] Raymond Doudlah,[1] and Ting-Yu Chang[2]

[1]Department of Neuroscience, School of Medicine and Public Health, University of Wisconsin–Madison, Madison, Wisconsin, USA; email: ari.rosenberg@wisc.edu

[2]School of Medicine, National Defense Medical Center, Taipei, Taiwan

Annu. Rev. Vis. Sci. 2023. 9:337–59

First published as a Review in Advance on March 21, 2023

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-111022-123857

Keywords

3D vision, stereopsis, perspective, depth, orientation, motion

Abstract

The visual system must reconstruct the dynamic, three-dimensional (3D) world from ambiguous two-dimensional (2D) retinal images. In this review, we synthesize current literature on how the visual system of nonhuman primates performs this transformation through multiple channels within the classically defined dorsal (where) and ventral (what) pathways. Each of these channels is specialized for processing different 3D features (e.g., the shape, orientation, or motion of objects, or the larger scene structure). Despite the common goal of 3D reconstruction, neurocomputational differences between the channels impose distinct information-limiting constraints on perception. Convergent evidence further points to the little-studied area V3A as a potential branchpoint from which multiple 3D-fugal processing channels diverge. We speculate that the expansion of V3A in humans may have supported the emergence of advanced 3D spatial reasoning skills. Lastly, we discuss future directions for exploring 3D information transmission across brain areas and experimental approaches that can further advance the understanding of 3D vision.

INTRODUCTION

Orbital convergence: the degree to which the orbits of the left and right eyes point in the same direction

Visual perception is compellingly three dimensional (3D). However, this perception is far removed from what our eyes sense because retinal images are two-dimensional (2D) and ambiguous to the veridical scene structure (Howard & Rogers 1995). Visually guided behaviors such as manipulating objects and navigating through volatile environments require that retinal images be transformed into ecologically relevant representations of 3D objects and scenes. In this review, we synthesize current literature on the cortical substrates and processing channels in nonhuman primates that support the 2D-to-3D visual transformation.

Comparative analyses across the primate order indicate that orbital convergence correlates with the size of the visual brain (Barton 2004). This relationship may reflect that evolutionary pressure to process 3D information was a driving factor in primate brain expansion, as well as a likely antecedent to human analytical abilities and the creation of tools. Indeed, 3D spatial reasoning skills predict performance in STEM (science, technology, engineering, and mathematics) fields (Uttal & Cohen 2012), and deficits in visuospatial processing are indicative of certain neurodevelopmental disorders (Banker et al. 2020). Such observations suggest that understanding the neurocomputational basis of 3D vision may yield broad insights into human cognitive capabilities and neurological conditions (Rosenberg et al. 2015).

The primate visual system is classically divided into dorsal (where) and ventral (what) pathways (Mishkin et al. 1983, Pohl 1973). In this review, we discuss evidence for parallel channels within the dorsal visual pathway that are specialized for processing different 3D features, including the static pose and motion of objects. Behavioral evidence suggests that neurocomputational differences between these channels may impose distinct information-limiting constraints on 3D perception that have yet to be explored at the neuronal level. We further discuss how 3D object and scene processing may likewise be supported by different channels within the ventral pathway. We conclude by positing a connection between the expansion of little-studied area V3A and the evolution of 3D spatial reasoning skills, as well as discussing the need for theoretically motivated, standardized experimental designs to clarify area-specific functions and the contributions of cross-channel interactions to 3D vision. This review thus explores parallel and hierarchical transformations that support 3D representations of a world that we cannot directly sense but with which we must nevertheless interact.

STATIC AND DYNAMIC THREE-DIMENSIONAL VISUAL CUES

The 3D position, orientation, and motion of objects are signaled by multiple visual signals including perspective and stereoscopic cues. Perspective and other monocular cues such as shading (Koenderink & van Doorn 2004, Todd & Mingolla 1983), occlusion (Harris & Wilcox 2009, Shimojo & Nakayama 1990), blur (Held et al. 2010, Mather 1996), and dynamic signals resulting from eye movements and self-motion (Kim et al. 2016, Rogers 1993) are available to each eye alone. Stereoscopic cues are created through binocular computations. In this section, we introduce perspective and stereoscopic cues that have been widely used to study 3D vision and discuss functional considerations for interpreting neuronal responses to such stimuli.

Static Perspective Cues

As a consequence of projective geometry, a 3D scene is sensed as a pair of 2D retinal images. Although the resulting perspective cues are ambiguous to the veridical scene, as revealed by perspective-based illusions, they can nevertheless be used to infer 3D structure (Knill 1998, Stevens 1981). For example, if a set of stationary elements in the world are uniformly spaced and sized, then elements closer to the observer will be larger (a scaling cue) and further apart (a density

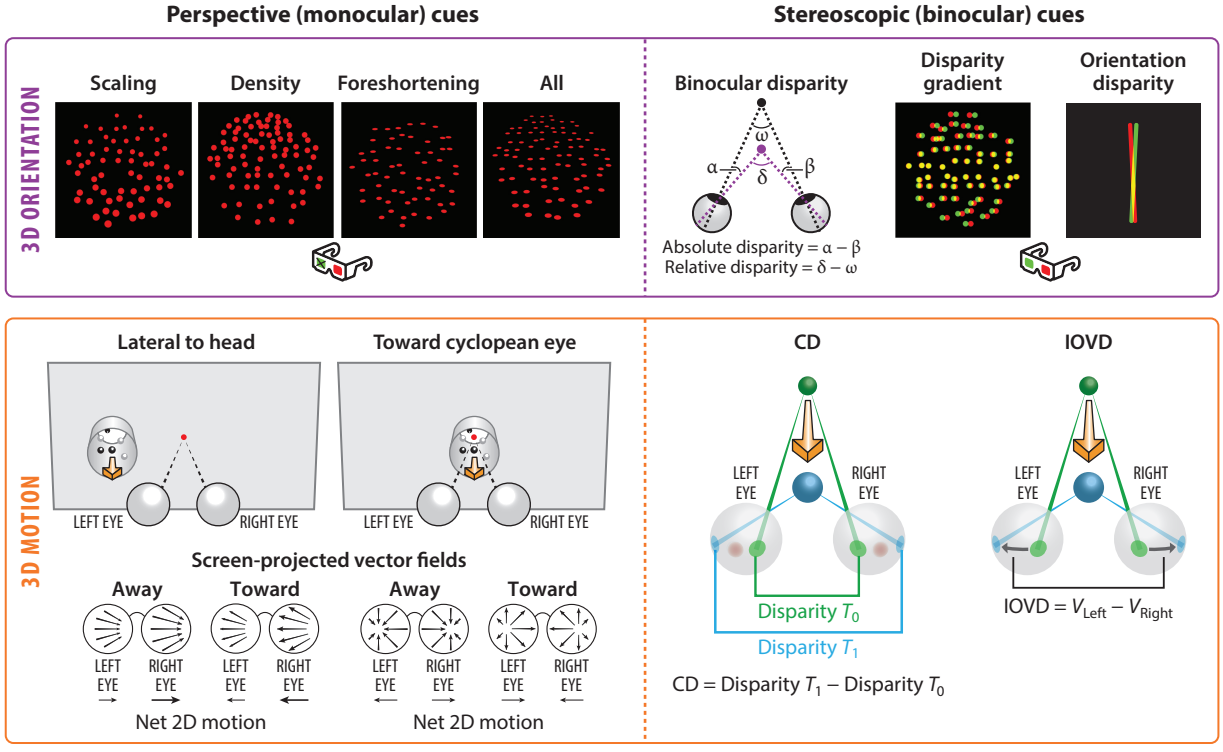

Figure 1

Perspective (*left column*) and stereoscopic (*right column*) cues supporting the perception of three-dimensional (3D) orientation (*top row*) and 3D motion (*bottom row*). (*Top left*) An object's 3D orientation is signaled by multiple perspective cues including scaling, density, and foreshortening (shown for a bottom-near plane with uniformly spaced dot elements in the world). Each cue is shown in isolation and together (all) as seen by the left eye. (*Top right*) Binocular disparity signals are quintessential stereoscopic depth cues. Provided that fixation is on the black point, the purple point has a near absolute disparity ($\alpha - \beta$; *left*). The relative disparity between the black and purple points, $\delta - \omega$, is independent of where the eyes are fixated. An object's 3D orientation is signaled by a disparity gradient (shown for a bottom-near plane; *middle*) as well as an orientation disparity (shown for a bottom-near bar; *right*). Examples are illustrated as red–green anaglyphs. (*Bottom left*) Optic flow is a prominent perspective cue to 3D motion. Patterns of optic flow differ in the two eyes due to their horizontal offset and depend on the object's 3D motion trajectory and visual field location. Objects lateral to the head and moving in depth produce larger velocities in the contralateral than the ipsilateral eye (shown as screen-projected vector fields). In contrast, motion toward or away from the cyclopean eye produces optic flow patterns with equal and opposite net two-dimensional (2D) retinal motions. (*Bottom right*) Two stereoscopic cues signaling 3D object motion. A dot begins (time T_0; *green*) at a far disparity and ends (time T_1; *blue*) at a near disparity (red dots correspond to the foveae). The difference in disparities at the two time points provides a changing disparity (CD) cue to 3D motion (*left*). The same motion also produces distinct left- (V_{Left}) and right-eye (V_{Right}) retinal velocities (*black arrows*; *right*). The difference between these velocities provides an interocular velocity difference (IOVD) cue to 3D motion.

cue) in the retinal images (**Figure 1**, *top left*). Likewise, the shape of the elements in the retinal images will depend on their orientations relative to each eye (a foreshortening cue). Importantly, the perspective cues sensed by the left and right eyes differ because the eyes are horizontally offset.

Dynamic Perspective Cues

For a moving object, projective geometry results in patterns of retinal motion called optic flow (Gibson 1947, Longuet-Higgins & Prazdny 1980). An object's 3D motion is further signaled by

changes in the retinal size, density, and shape of visual elements over time. Notably, the optic flow patterns sensed by the two eyes can dramatically differ depending on the object's 3D trajectory (Cormack et al. 2017; Thompson et al. 2019, 2021). For example, an object that is located lateral to the head and moving in depth produces retinal-motion signals with the same net direction but different speeds in the two eyes (**Figure 1**, *bottom left*). In contrast, an object moving toward (or away) from the cyclopean eye produces equal and opposite net 2D retinal-motion signals in the two eyes (**Figure 1**, *bottom left*).

Cyclopean eye: the midpoint between the two eyes

Absolute disparity: the angular difference between left and right retinal image positions of a point in 3D space relative to the foveae

Relative disparity: the difference in absolute disparities of two points in 3D space

Vergence: the angle between the visual axes of the two eyes at the point at which they converge in 3D space

Neuronal responses to 3D perspective cues have been measured either by binocularly presenting a stimulus rendered for the cyclopean eye (eliminating naturally occurring eye-specific information) or by monocularly presenting a stimulus appropriately rendered for that eye (preserving the natural geometry). For binocularly presented cyclopean eye stimuli, it is important to consider that the perspective cues are not properly rendered for either eye. Such stimuli also contain stereoscopic cues (because both eyes are visually stimulated) signaling a single depth, and therefore have a cue conflict if the perspective cues signal multiple depths. Within the caudal intraparietal (CIP) area, 3D orientation–selective neurons are sensitive to the difference in these two approaches to rendering perspective cue stimuli (Rosenberg & Angelaki 2014b). Behavioral studies further show that eye-specific perspective cues to 3D motion can support significantly different levels of visual sensitivity (Thompson et al. 2019, 2021). For these reasons, we favor the monocular presentation of perspective cue stimuli with proper eye-specific geometry (as if one eye were closed).

Stereoscopic Cues

Because the two retinal images cast by a 3D object systematically differ, comparisons of them allow for the computation of stereoscopic (binocular) cues that support 3D perception (Cumming & DeAngelis 2001, Howard & Rogers 1995). Most notably, absolute disparity, the difference in the two corresponding retinal image positions of a point in 3D space, signals local depth information relative to the depth of fixation (**Figure 1**, *top right*). Absolute disparity provides a quintessential signal for performing stereoscopic computations. This includes the calculation of relative disparity, upon which stereoscopic depth perception largely depends. Importantly, unlike absolute disparity, relative disparity signals are independent of where the eyes look (i.e., vergence) (Parker 2007). Higher-order spatial variations such as disparity gradients further signal an object's 3D shape or orientation. Orientation disparity, the difference in the two retinal image orientations of an object, also signals 3D orientation (Greenwald & Knill 2009). Two dynamic stereoscopic cues further support the perception of motion in depth: (*a*) changes in disparity over time (Cumming & Parker 1994, Czuba et al. 2011) and (*b*) interocular velocity differences (Brooks 2002, Rokers et al. 2008) (**Figure 1**, *bottom right*).

Stereoscopic cue stimuli are typically rendered using orthographic projection to create random dot patterns whose elements have a constant retinal size, density, and shape (**Figure 1**, *top right*). For 3D motion, the resulting retinal signals reduce to 2D horizontal translations whose left or right directions and speeds depend on the 3D trajectory. Unfortunately, 3D stereoscopic cue stimuli contain inescapable cue conflicts. For example, when stereoscopic cues are used to define a plane oriented in depth, there are still perspective cues (i.e., same-sized, evenly spaced, and isotropic dot elements) that instead signal a frontoparallel plane. Although such conflicts cannot be eliminated, 3D discrimination tasks are ideally designed such that the perspective cues are uninformative, and control experiments have suggested that the impact on perception is minimal, at least for well-trained observers (Chang et al. 2020b, Hillis et al. 2004, Thompson et al. 2021).

A DORSAL CHANNEL REPRESENTING THREE-DIMENSIONAL OBJECT POSE

3D pose: the position and orientation of an object in 3D space

Cue integration: the process of integrating different sensory estimates to improve the precision of perception

The 3D pose of a rigid-body object has six degrees of freedom. Three translational degrees of freedom specify its position, and three rotational degrees of freedom specify its orientation. In contrast, retinal images are 2D and therefore possess only two translational and one rotational degrees of freedom that confound 3D position and orientation information. To reconstruct the missing degrees of freedom, the visual system must disambiguate this information. In this section, we review evidence for a channel within the dorsal pathway that computes 3D object pose from 2D retinal images.

As illustrated in **Figure 2**, the stereoscopic processing of depth begins with a subset of primary visual cortex (V1) neurons selective for absolute disparity (Cumming & Parker 1999). A transformation from absolute to relative disparity that could support depth perception occurs, at least partially, at the level of V2 (Thomas et al. 2002). In addition, V2 may contain neurons selective for gradients of absolute disparity (Du & Ts'o 2013). Downstream of V2, anatomical (Nakamura et al. 2001) and neuroimaging (Tsao et al. 2003, Van Dromme et al. 2016) evidence has implicated two regions bridging the parieto-occipital junction, areas V3A and CIP, in 3D vision.

The computational role of V3A has been challenging to discern due to conflicting findings. Consistent with low-level image processing, some studies reported spatiotemporal filtering properties qualitatively similar to those in V1 (Gaska et al. 1987, 1988), prominent absolute disparity but limited relative disparity processing (Anzai et al. 2011), and 2D direction selectivity (Nakhla et al. 2021). Other work suggested that V3A supports higher-level visual functions. In particular, multiple studies found V3A neurons whose visual selectivity was modified by extraretinal signals, consistent with the creation of allocentric object representations (Galletti & Battaglini 1989; Galletti et al. 1990; Nakamura & Colby 2000, 2002; Sauvan & Peterhans 1999), and some V3A neurons show 3D orientation selectivity (Doudlah et al. 2022, Elmore et al. 2019). These contrasting results collectively suggest that V3A contains heterogeneous visual representations and indicate that the area requires further study.

By comparison, a role for CIP in 3D vision is well-established. Early work discovered CIP neurons selective for both stereoscopic and perspective cues to 3D planar orientation and showed that the preferences were often similar for the two cues (Tsutsui et al. 2001, 2002). Subsequent work found that CIP neurons that integrated those cues tended to weight perspective cues more heavily if they had larger slant preferences, consistent with reliability-dependent cue integration (Rosenberg & Angelaki 2014b). While hierarchical transformations of stereoscopic signals that could support 3D orientation representations in CIP are relatively well-documented (**Figure 2**), corresponding transformations of perspective signals have yet to be identified. Neurons in CIP also carry memory- and choice-related activity during 3D discrimination tasks (Chang et al. 2020a, Elmore et al. 2019, Tsutsui et al. 2003), although causal evidence linking the area to 3D pose perception remains rather limited (Tsutsui et al. 2001, Van Dromme et al. 2016). Intriguingly, at rolled head/body orientations, gravitational information provided by vestibular/proprioceptive signals underlies changes in the 3D orientation preferences and response amplitudes of CIP neurons (Rosenberg & Angelaki 2014a). These changes can support a flexible neural code for representing 3D object orientation in either egocentric or allocentric reference frames, depending on task demands. Together, these studies showed that CIP contains high-level 3D representations, but until recently it was unclear where 3D pose representations emerge in the visual hierarchy.

To provide a theoretical rationale for distinguishing 3D pose tuning from lower-level feature selectivity, we briefly introduce the geometry of 3D pose (Cowan & Chang 2005, Cowan et al. 2002). The 3D position of a rigid-body object is parameterized by a 3-vector, $p \in \mathbb{R}^3$ (Euclidean

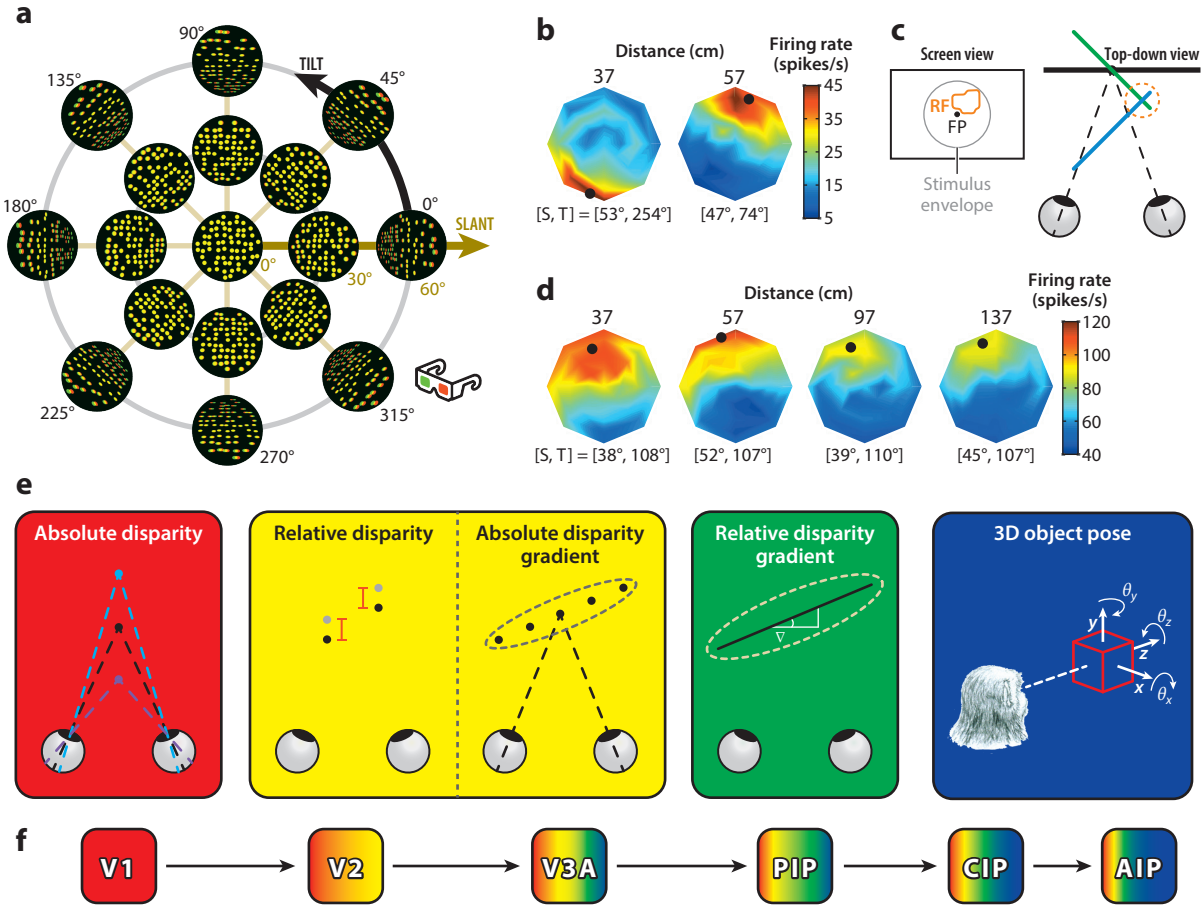

Figure 2

Neuronal representations supporting 3D object pose. (*a*) Slant–tilt coordinates for 3D planar orientation. Slant (radial axis; *gold*) specifies the orientation in depth. Tilt (angular axis; *black*) specifies which side of the plane is nearest to the observer. The example planes are rendered with perspective and stereoscopic cues as red–green anaglyphs. (*b*) A V3A neuron with orientation tuning at 37 and 57 cm (fixation at 57 cm). Heat maps plot the neuron's responses in slant–tilt coordinates. Black dots mark the preferred slant (S) and tilt (T). Note that the slant preferences were similar, but the tilt preferences were 180° apart at 37 and 57 cm. (*c*) Schematic illustrating how absolute disparity selectivity can account for this tuning. Planes at 37 cm (*blue*) and 57 cm (*green*) with equal slants and tilts 180° apart intersect within the orange encircled region, which corresponds to near disparities. The inset shows the neuron's RF (*orange contour*), the FP (*black dot*), and the stimulus envelope (*gray circle*). The neuron's RF position and orientation preferences at 37 and 57 cm together imply a preference for near disparities. (*d*) A CIP neuron with orientation tuning at each distance (fixation at 57 cm). The response amplitude varied with distance, but the shape of the orientation tuning curve was similar, implying 3D pose tuning. (*e*) Hierarchical transformations of stereoscopic signals culminating in the representation of 3D object pose. First, absolute disparities are computed (*red box*). Provided that fixation is on the black point, the purple and blue points have near and far disparities, respectively. Second, relative disparity (*yellow box*, *left*) and absolute disparity gradient (*yellow box*, *right*) representations are computed. Third, relative disparity gradients are computed such that the preferred gradients (∇) are depth tolerant (*green box*). Finally, 3D object pose representations are achieved with multiplicatively separable tuning for 3D position (x, y, z) and orientation (θ_x, θ_y, θ_z) (*blue box*). (*f*) Dorsal channel computing 3D object pose from 2D retinal images. Arrows indicate feedforward projections. Areas are approximately color-coded according to the representations in panel *e* that they contain. Abbreviations: 2D, two-dimensional; 3D, three-dimensional; AIP, anterior intraparietal; CIP, caudal intraparietal; FP, fixation point; PIP, posterior intraparietal; RF, receptive field.

space), where each element specifies a location along one dimension. The object's 3D orientation is likewise parameterized by three 3 × 3 rotation matrices, denoted by the Special Orthogonal Group, SO(3), each of which specifies a rotation about an axis. Geometrically, SO(3) corresponds to a 3-sphere with antipodal symmetry, which fundamentally constrains the shape of neuronal 3D orientation tuning curves (Rosenberg et al. 2013). In turn, an object's 3D pose is defined by the Cartesian product of $\mathbb{R}^3$ and SO(3), the Special Euclidean Group: $SE(3) = \mathbb{R}^3 \times SO(3)$. The visual system presumably transforms 2D retinal images into 3D pose representations through a series of hierarchical steps. As this transformation progresses, the geometry of SE(3) predicts that neuronal responses should show increasing multiplicative separability over position and orientation.

A practical challenge to assessing 3D pose tuning is the high dimensionality of SE(3). One way to circumvent this challenge is to make the problem more tractable by focusing on critical degrees of freedom with judiciously chosen stimuli. For example, varying only the distance of planes or bars with texture patterns that are rotationally invariant about one axis effectively reduces the degrees of freedom from six to three (one translational and two rotational). Distinguishing 3D pose from lower-level feature selectivity can be further facilitated by holding the fixation distance constant while varying the stimulus distance, since this inflates differences in the low-level absolute disparity signals across stimuli. Using this strategy, Chang et al. (2020a) and Doudlah et al. (2022) recently compared the visual representations found in V3A and CIP. Specifically, they examined how the 3D orientation tuning curves of single neurons depended on the distance of planar surfaces defined by random dots with perspective and stereoscopic cues (**Figure 2*a***). The representations in both areas were surprisingly heterogeneous. For some neurons, especially in V3A, the orientation tuning curve shape greatly differed with distance (**Figure 2*b***). Because the perspective cues defining the planes did not depend on distance (only on orientation), this finding implied selectivity for low-level absolute disparity signals (**Figure 2*c***). Other neurons showed orientation tuning at a single distance only, consistent with an intermediate representation of absolute disparity gradients, perhaps similar to some V2 (Du & Ts'o 2013) and middle temporal (MT) area neurons (Nguyenkim & DeAngelis 2003). The orientation tuning of the remaining neurons was highly tolerant to distance, especially in CIP, implying selectivity for higher-level relative disparity gradients. A quantitative assessment of the multiplicative separability of the joint tuning for position and orientation further indicated that, for many of those neurons, the amplitude but not the shape of the orientation tuning curve depended on distance (**Figure 2*d***), consistent with the $\mathbb{R}^3 \times SO(3)$ geometry. This representation implies 3D pose tuning and has notable decoding benefits, since either position or orientation can be readily decoded after marginalizing over the other variable. Highlighting the hierarchical nature of the 2D-to-3D visual transformation, the multiplicative separability of 3D pose tuning curves significantly increased from V3A to CIP (Doudlah et al. 2022).

While functional and anatomical data support a V3A → CIP hierarchy in the computation of 3D pose, effective connectivity work has also implicated the posterior intraparietal (PIP) area (Van Dromme et al. 2016). Specifically, electrical microstimulation of PIP revealed effective connectivity with V3A and CIP. However, when CIP was stimulated, effective connectivity was observed in PIP but not V3A. These findings suggest that PIP may reside hierarchically between V3A and CIP. Neuronal recordings with stereoscopic cue stimuli further indicated that PIP contains heterogeneous visual representations, similar to V3A and CIP (Alizadeh et al. 2018a). However, unlike in CIP, which contains a roughly uniform representation of all 3D orientations (Chang et al. 2020a, Elmore et al. 2019, Rosenberg et al. 2013), there was an overrepresentation of frontoparallel-like stimuli in PIP. One possibility to explore is that representations of relative disparity gradients that begin to emerge in V3A are solidified in PIP before 3D pose tuning is more widely achieved in

CIP. It is also likely that 3D visual representations are further transformed downstream of CIP in the anterior intraparietal (AIP) area, as well as in premotor cortex, to estimate 3D shape (Durand et al. 2007, Srivastava et al. 2009, Theys et al. 2012) and support prehensile object manipulations (Murata et al. 2000).

Correspondence problem: the problem of identifying matching features in the two retinal images that correspond to a common feature in the world

Existing evidence thus suggests that stereoscopic signals are hierarchically transformed from low-level absolute disparities to high-level 3D object pose representations (**Figure 2*e***) along a channel within the dorsal pathway, as illustrated in **Figure 2*f***. Still, much remains to be discovered about how those transformations are implemented. For instance, the extent to which V3A, PIP, and CIP neurons can discern which 2D image features in the two eyes belong to the same object and reject false matches (i.e., solve the correspondence problem) remains unknown. It is also unclear how perspective signals are hierarchically transformed along this pathway and at which stages cue integration supports the creation of robust 3D visual representations.

A DORSAL CHANNEL REPRESENTING THREE-DIMENSIONAL OBJECT MOTION

While much is known about the processing of 2D retinal motion (Born & Bradley 2005, Pasternak & Tadin 2020), the extent to which that knowledge can inform the ecologically relevant domain of 3D object-motion processing is limited because the computations required to estimate 3D motion are distinct. In particular, objects moving in depth produce more complex retinal-motion patterns (optic flow) (Longuet-Higgins & Prazdny 1980) and slower retinal speeds than are typically used to study 2D motion processing (Cooper et al. 2016, Cormack et al. 2017). For example, an object that is 2 m from an observer and moving at 1.34 m/s (a walking pace) produces drastically different retinal speeds depending on whether the direction is perpendicular to (2D motion; approximately 34°/s) or toward (3D motion; approximately 1.86°/s) the observer. In this section, we review the current understanding of a dorsal channel that computes 3D object motion from 2D retinal motion.

As illustrated in **Figure 3**, the cortical processing of motion begins with a subset of V1 neurons selective for the 2D direction of retinal motion produced by simple patterns such as bars, dots, or gratings (De Valois et al. 1982, Orban et al. 1986). However, few V1 neurons are selective for the perceived 2D motion of even slightly more complex patterns created by the superposition of two gratings (Movshon et al. 1985). Direction-selective V1 neurons project to MT (Movshon & Newsome 1996), where 2D pattern motion selectivity emerges. Neurons in MT further show tuning for multiple features that can support 3D motion processing, including speed (Nover et al. 2005), binocular disparity (DeAngelis & Uka 2003), and motion parallax (Nadler et al. 2008). However, the role of MT in 3D motion processing remains controversial.

Two earlier studies reported a small number of MT neurons with opposite 2D direction preferences in the two eyes (Albright et al. 1984, Zeki 1974), which could conceivably make them selective for 3D motion toward or away from the observer (Poggio & Talbot 1981, Sabatini & Solari 2004). However, other work concluded that the apparent 3D motion tuning of most MT neurons was attributable to 2D motion and static disparity selectivity, which, under certain circumstances, gave an impression of 3D selectivity (Maunsell & Van Essen 1983). Moreover, MT neurons typically respond to local 2D retinal motion signals when presented with large-field optic flow stimuli (Lagae et al. 1994, Nakhla et al. 2021). The possibility of 3D motion tuning acquired subsequent support from two studies that reported that approximately 58% to 70% of MT neurons were selective for the direction of stereoscopically defined 3D motion (Czuba et al. 2014, Sanada & DeAngelis 2014). That selectivity was initially attributed, at least in part, to nonlinear summation of monocular 2D motion signals. However, subsequent computational work indicated

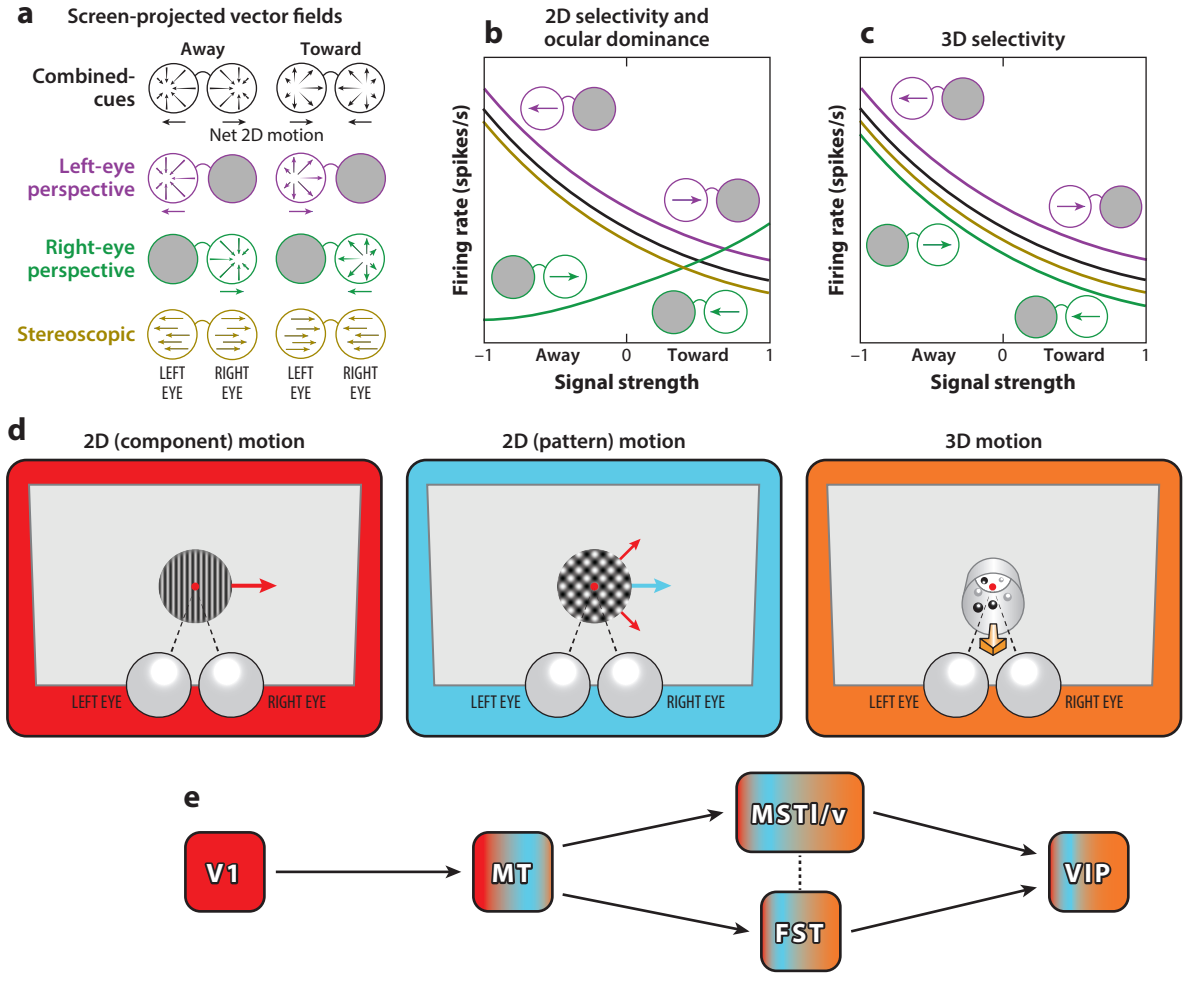

Figure 3

Neuronal representations supporting 3D object motion. (*a*) Screen-projected vector fields for 3D motion toward and away from the cyclopean eye for different cue conditions: perspective and stereoscopic cues together (combined-cues; *black*), left-eye perspective cues (*purple*), right-eye perspective cues (*green*), and stereoscopic cues (*gold*). Left- and right-eye perspective cue stimuli have distinct optic flow patterns with equal and opposite net 2D retinal motions. Stereoscopic cue stimuli have equal and opposite horizontal translations in the two eyes. (*b*) Expected responses to these stimuli for a neuron with 2D leftward motion selectivity and left eye ocular dominance. The strongest responses will be to away motion for left-eye perspective cue stimuli but to toward motion for right-eye perspective cue stimuli. Responses to stereoscopic cue and combined-cue stimuli will follow the dominant eye. Encircled arrows indicate the net 2D retinal motions for the perspective cue stimuli. (*c*) Expected responses for a neuron with 3D motion selectivity. Stimulus preferences will be tolerant to the cue condition. (*d*) Hierarchical transformations of motion signals culminating in the representation of 3D motion. First, the component of retinal motion (*red arrow*) signaled by a drifting grating is computed (*red box*). Second, pattern motion is computed from the component motions (*blue arrow* and *blue box*). Finally, 3D object motion representations are achieved (*orange box*). (*e*) Dorsal channel computing 3D object motion from 2D retinal motion. Arrows indicate feedforward projections. The relationship between the FST and MSTl/v in 3D motion processing is currently unknown (*dotted line*). Areas are approximately color coded according to the representations in panel *d* that they contain. Abbreviations: 2D, two-dimensional; 3D, three-dimensional; FST, fundus of the superior temporal sulcus; MSTl/v, lateral/ventral medial superior temporal; MT, middle temporal; VIP, ventral intraparietal.

Ocular dominance: the extent to which a binocularly driven neuron responds more strongly when a particular eye is visually stimulated

that the responses could largely be explained by a linear summation model that incorporated monocular 2D direction selectivity and ocular dominance (Bonnen et al. 2020). This was proposed to be the basis of a noncanonical mechanism for representing 3D motion, but a more parsimonious explanation is that 3D motion is computed downstream of MT. Indeed, if 2D direction selectivity and ocular dominance provide the basis for 3D motion coding, then 2D direction–selective V1 neurons (or, for that matter, retinal ganglion cells) should also be classified as 3D selective. After all, they too will respond preferentially to 3D motions that result in the dominant eye seeing the preferred 2D motion. Thus, more conservative criteria are required to classify a neuron as 3D motion selective.

One possibility is to test if a neuron's preference for motion toward or away from the cyclopean eye is invariant to the defining cue(s). Specifically, 2D versus 3D selectivity can be distinguished by comparing responses to these motions defined by left-eye perspective cues, right-eye perspective cues, stereoscopic cues, and all three cues combined (**Figure 3*a–c***). In this way, Thompson et al. (2022) recently assessed the 3D motion selectivity of MT neurons using small-field stimuli. Across all cue conditions, 2D direction selectivity and ocular dominance accounted for the responses of most MT neurons (Thompson et al. 2022). Most notably, the direction preferences for left- and right-eye perspective cue stimuli were generally opposite, implying that the responses reflected the net direction of 2D retinal motion rather than the 3D object motion (**Figure 3*b***). These results suggest that MT is largely specialized for processing 2D retinal motion. Where might the output of MT be used to compute 3D object motion?

Three downstream targets of MT include the dorsal medial superior temporal (MSTd) area, lateral/ventral medial superior temporal (MSTl/v) area, and fundus of the superior temporal sulcus (FST) (Ungerleider & Desimone 1986). Area MSTd represents self-motion through the multisensory processing of large-field optic flow patterns and vestibular/proprioceptive signals (Angelaki et al. 2011). In contrast, MSTl/v is a likely candidate for object-motion processing because the neurons are also selective for optic flow but have smaller visual receptive fields (Eifuku & Wurtz 1998, Tanaka et al. 1993). In addition, many MSTl/v neurons carry smooth pursuit and/or vestibular/proprioceptive signals, which could support reference frame transformations for representing object motion in head, body, or world coordinates (Ilg et al. 2004, Sasaki et al. 2020). However, clarifying the role of MSTl/v in 3D object-motion processing will require further work using small-field stimuli, assessing the extent to which the correspondence problem has been resolved, and testing if 3D direction selectivity is cue invariant.

A recent neuroimaging study implicated FST in 3D motion processing (Héjja-Brichard et al. 2020). Specifically, this work found greater sensitivity to stereoscopically defined 3D motion in FST than in MT, consistent with earlier findings of FST neurons that were selective for opposite directions of 2D motion (Mysore et al. 2010, Rosenberg et al. 2008). Further support for this possibility comes from Thompson et al. (2022), who found that the preferences of FST neurons for motion toward or away from the cyclopean eye were often cue invariant, in stark contrast to MT. Most notably, the preferences for left- and right-eye perspective cue stimuli were generally the same (despite the fact that the net 2D retinal motion directions were opposite), implying a robust representation of 3D motion (**Figure 3*c***). Given these findings, it will be important for future work to evaluate the extent to which FST neurons have resolved the correspondence problem, determine if they carry extraretinal signals (Bogadhi et al. 2021), and assess the reference frame(s) in which they represent 3D motion. Clarification of the larger cortical network that supports 3D object-motion processing will also require determining the functional and potentially hierarchical relationship between FST and MSTl/v.

Beyond representing an object's motion in eye or head coordinates, it is often necessary to determine its motion relative to external (world) references. Extraretinal signals that may support

this reference frame transformation are found as early as MT (Kim et al. 2022) and MSTl/v (Ilg et al. 2004), at least for 2D motion. However, it is at the level of the ventral intraparietal (VIP) area, which contains a rotation-invariant representation of self-motion (Sunkara et al. 2015, 2016), that a representation of object motion in world coordinates is likely achieved. Specifically, recent work found that VIP neurons utilize 2D object motion and self-motion signals to flexibly represent an object's motion in either egocentric or allocentric coordinates, depending on task demands (Sasaki et al. 2020). It will be important for future work to test if this flexible representation generalizes to 3D object-motion trajectories and if VIP causally contributes to object-motion perception in egocentric and/or allocentric reference frames.

Existing evidence thus suggests that 2D retinal-motion signals are transformed into ecologically relevant 3D object-motion representations (**Figure 3*d***) along a channel within the dorsal pathway, as illustrated in **Figure 3*e***. Still, much remains to be discovered about the constituent roles of these areas and the functional interactions between them in the processing of 3D object motion.

COMPUTATIONAL DIFFERENCES IN THE PROCESSING OF THREE-DIMENSIONAL POSE AND MOTION

Inferences about an object's 3D pose and motion are limited by environmental and neural noise but can be improved by integrating estimates based on different cues (Clark & Yuille 1990, Knill & Richards 1996). For a linear integration process, the result is statistically optimal if the individual estimates are proportionally weighted by their reliabilities. Such models generally assume that different sensory signals support statistically independent estimates. This assumption is likely most valid when the signals are detected by different sensory organs and initially represented by distinct neural populations with predominantly different noise sources (Ma et al. 2006, Seilheimer et al. 2014), a possibility supported by the reliability-dependent reweighting of visual and vestibular cues to self-motion (Fetsch et al. 2011). However, the assumption may break down for within-modality cue integration, especially if the same neurons create multiple estimates based on different signals. Such multiplexed processing could impose information-limiting constraints on the perceptual gains associated with cue integration (Oruç et al. 2003), as was recently tested for 3D orientation and motion perception, with important implications regarding the underlying neural computations (Chang et al. 2020b, Fulvio et al. 2020, Thompson et al. 2021).

Surface orientation perception studies with humans provided an early demonstration of cue integration (Hillis et al. 2004, Knill & Saunders 2003) that was recently corroborated by work with macaques (Chang et al. 2020b). These studies found that, when observers judged the 3D orientation of a plane, the perception of combined-cue stimuli was well-described by the linear integration of perspective and stereoscopic cues weighted by their reliabilities. In particular, the reliability of perspective cues increases with slant, and the reliability of stereoscopic cues decreases with distance. Correspondingly, the relative contribution of perspective cues to the perceived orientation of combined-cue stimuli increased as either the slant or distance increased. This reliability-dependent integration of perspective and stereoscopic cues may be supported by CIP (Rosenberg & Angelaki 2014b). Furthermore, the perception of combined-cue stimuli was consistent with the visual system forming (*a*) a stereoscopic cue estimate of the plane's orientation and (*b*) a single perspective cue estimate (as opposed to separate perspective estimates for each eye). This finding was explained by a model in which orientation estimates were created by two independent neuronal populations, with one relying on stereoscopic cues and the other relying on perspective cues indiscriminately pooled from the two eyes. Most notably, forming a single perspective cue estimate imposed a steep information-limiting constraint on cue integration such that 3D orientation perception was half as precise as theoretically possible (Chang et al. 2020b).

Divisive normalization (DN): a canonical neural computation that divides the response of an individual neuron by a weighted sum of population activity

The generality of this information-limiting constraint on 3D perception was explored in parallel studies of 3D motion. For 3D motion, the reliabilities of perspective and stereoscopic cues also depend on viewing geometry. For example, an object that is lateral to the head and moving in depth produces less reliable perspective cues in the ipsilateral eye than in the contralateral eye due to differences in the retinal velocities (**Figure 1**, *bottom left*). With such unbalanced signals, the perception of combined-cue 3D motion was found to approximately follow a winner-take-all strategy that was dominated by the more reliable perspective cue (provided by the contralateral eye), with little to no contribution of the other eye's perspective cues or stereoscopic cues (Thompson et al. 2019). In contrast, motion toward or away from the cyclopean eye produces balanced left- and right-eye perspective cues with similar reliabilities (**Figure 1**, *bottom left*). With these signals balanced, cue integration was apparent, and the information loss was much less than was observed for 3D orientation perception. Specifically, the perception of combined-cue 3D motion was more precise than expected based on the integration of a stereoscopic cue estimate and a single perspective cue estimate (Fulvio et al. 2020, Thompson et al. 2021). Instead, the results were explained by a model that integrated left- and right-eye perspective cue estimates that were partially to fully independent (depending on the observer) and an independent stereoscopic cue estimate. These results thus suggest that 3D motion processing makes more effective use of available sensory information than does 3D orientation processing. What might account for this difference?

Simulations performed by Chang and colleagues (2020b) suggest that it could reflect a combination of quadratic nonlinearities acting on the neuronal responses to the individual cues and a difference in the strength of divisive normalization (DN) during the integration of left- and right-eye perspective cues (**Figure 4**). If the DN signal was appropriately balanced relative to the effects of the quadratic nonlinearities, then cue integration would be linear, and the neuronal responses to the individual perspective cues would be additive (Ma et al. 2006, Qamar et al. 2013) (**Figure 4*a*,*b***). In this case, the two perspective cues would make independent contributions to perception, as has been found for some observers judging 3D motion (**Figure 4*c***). However, a sufficiently strong DN signal would result in a complete dependency between the representations of the two perspective cues. In this case, the neuronal responses to the individual cues would average when the cues have similar reliabilities (**Figure 4*a*,*b***). This subadditivity in response summation would reduce the perceptual gains associated with cue integration, as has been found for observers judging 3D orientation (**Figure 4*c***). Lastly, DN signals of intermediate strengths would result in partially dependent left- and right-eye perspective cue representations and correspondingly intermediate perceptual gains (**Figure 4*c***, *inset*), as has been found for some observers judging 3D motion. Differences in the effectiveness of cue integration may therefore reflect stronger DN during the integration of left- and right-eye perspective cues in the pose channel than in the motion channel. If neuronal recording studies confirm this difference, then it will be important to test whether it reflects a fixed network property or depends on the stimulus or task demands.

These studies collectively suggest that perspective and stereoscopic cues support effectively independent sensory estimates of both 3D orientation and motion, whereas the extent to which left- and right-eye perspective cues support separate estimates differs between the processing channels. This difference may have downstream consequences for behaviors that depend on an object's orientation and motion. For example, it suggests that observers will be worse at estimating a moving object's orientation than its trajectory, which could constrain the types of errors made in tasks such as catching a frisbee. Such considerations highlight the need for sensorimotor studies that require the simultaneous estimation of 3D object orientation and motion.

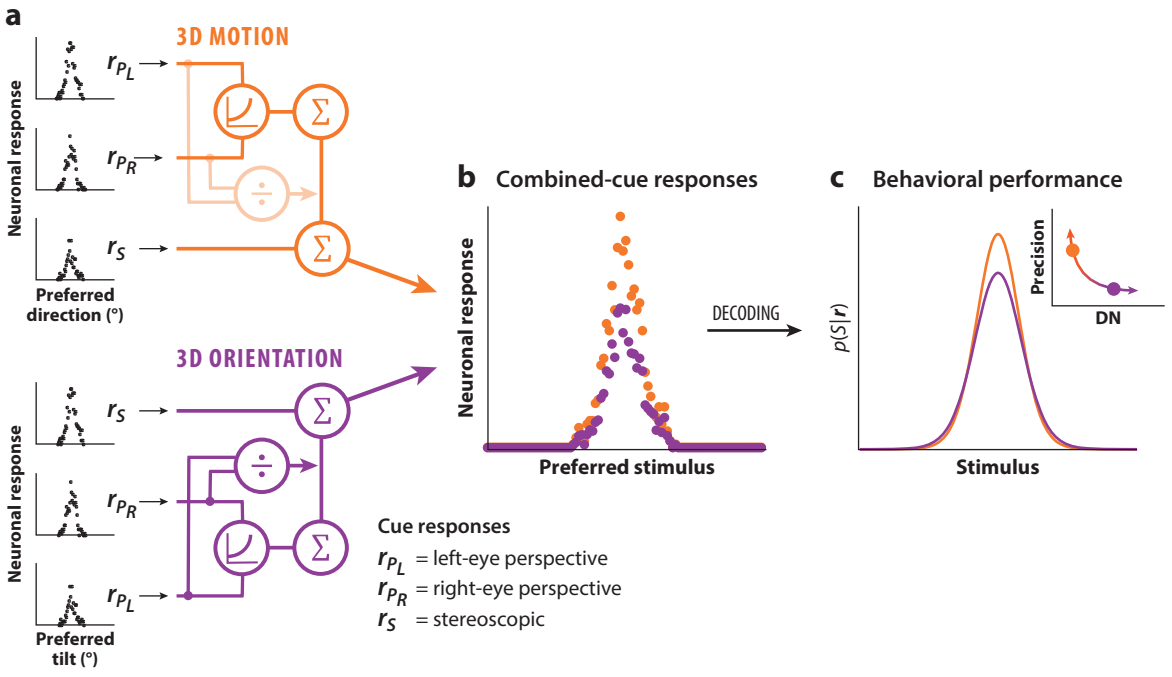

Figure 4

Computational differences in the neuronal processing of 3D motion and orientation, as suggested by perceptual studies. (*a*) Schematic architectures illustrating how the responses of populations of neurons representing left-eye perspective (r_{P_L}), right-eye perspective (r_{P_R}), and stereoscopic (r_S) cues are integrated during the processing of 3D motion (*top*; *orange*) and 3D orientation (*bottom*; *purple*). Black points show the responses of populations of hypothetical cue-selective neurons to a single stimulus (3D motion direction, *top*; planar tilt, *bottom*), ordered by the neurons' preferred stimulus. Responses to left- and right-eye perspective cues are passed through quadratic nonlinearities before they are integrated and divisively normalized. Depending on the strength of divisive normalization (DN) (illustrated as transparency), the contributions of the two perspective cues to the combined-cue responses can be additive (weaker DN, more effective) or subadditive (stronger DN, less effective). The perspective and stereoscopic cue responses are also integrated. (*b*) Differences in the strength of DN during the integration of the perspective cues result in more effective cue integration for 3D motion than for 3D orientation. Colored points show the responses of populations of hypothetical cue-integrating neurons to a single stimulus (motion or tilt), ordered by the neurons' preferred stimulus. (*c*) Posterior distributions, $p(S|\boldsymbol{r})$, describing the likelihood of stimulus S given combined-cue population response $\boldsymbol{r}$. The amplitude differences between the responses in panel *b* translate into a taller and narrower posterior over 3D motion than over 3D orientation. Perceptual estimates are correspondingly more precise (reliable) for 3D motion than for 3D orientation. The inset illustrates the inverse relationship between the precision of combined-cue perception and the strength of DN during the integration of the perspective cues.

VENTRAL PROCESSING OF THREE-DIMENSIONAL OBJECTS AND SCENES

The ventral pathway serves a critical role in object recognition. Although early work demonstrated this using real-world objects (Desimone et al. 1984, Kobatake & Tanaka 1994), subsequent studies largely leveraged the capabilities of 2D displays to investigate image-level computations. However, real-world objects and the contextual environments in which they reside are 3D, suggesting that depth information may contribute to object and scene processing (Humphrey & Khan 1992, Oliva & Torralba 2007). Consistent with this possibility, 3D representations are found within interconnected functional domains distributed throughout the ventral pathway (Premereur & Janssen 2020, Verhoef et al. 2016). In this section, we review evidence that the ventral pathway contains parallel channels that process 3D objects and scenes (**Figure 5**, *bottom right*).

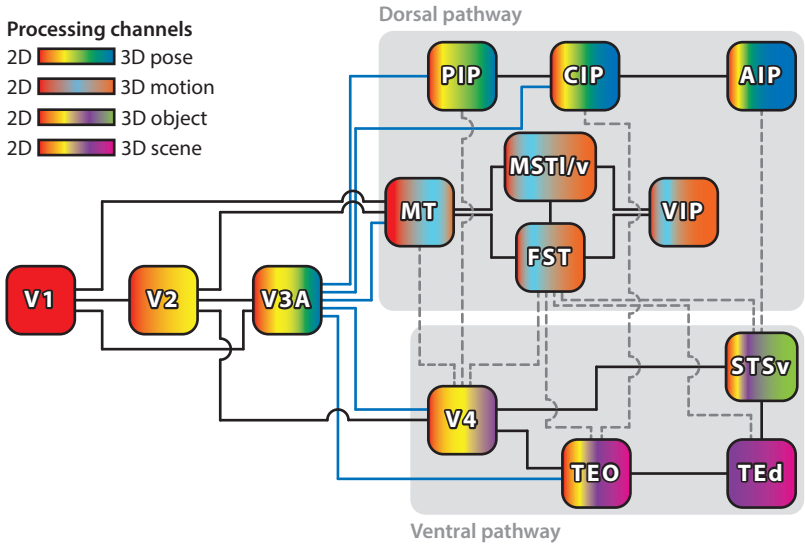

Figure 5

Processing channels within the dorsal and ventral pathways supporting 3D vision. Within each channel, visual signals are hierarchically transformed from ambiguous 2D retinal images into high-level representations of specific 3D features, as indicated by different color gradients. Note the heterogeneity of feature representations within most areas. Gray boxes demarcate the dorsal and ventral pathways. Within-pathway connections are indicated by black solid lines. Cross-pathway connections are indicated by gray dashed lines. Not all connections between the areas are shown. Growing evidence points to V3A as a nexus of 3D processing from which multiple 3D-fugal channels in the dorsal and ventral pathways originate (*solid blue lines*). Abbreviations: 2D, two-dimensional; 3D, three-dimensional; AIP, anterior intraparietal; CIP, caudal intraparietal; FST, fundus of the superior temporal sulcus; MSTl/v, lateral/ventral medial superior temporal; MT, middle temporal; PIP, posterior intraparietal; STSv, ventral bank of the superior temporal sulcus; TEO, posterior inferotemporal cortex; TEd, lateral aspect of TE; VIP, ventral intraparietal.

Visual representations from V2 and V3A are further processed in ventral area V4 (Felleman & Van Essen 1991, Ungerleider et al. 2008). Neurons in V4 show selectivity for features such as color (Kotake et al. 2009), 2D orientation (Hinkle & Connor 2002), and simple 2D shapes or contours (Hegdé & Van Essen 2007, Pasupathy & Connor 2001). They also contribute to stereoscopic depth processing. Compared to V1, the discernment of which 2D image features in the two eyes belong to the same object and rejection of false matches are more prevalent in V4, implying that V4 plays a role in solving the correspondence problem (Tanabe et al. 2004). Likewise, the proportion of neurons signaling relative, as opposed to absolute, disparity increases substantially from V2 to V4 (Hinkle & Connor 2001, Parker 2007, Umeda et al. 2007). Other work identified V4 neurons tuned for the 3D orientation of bars and perhaps planes (Hegdé & Van Essen 2005, Hinkle & Connor 2002). At least for bars, this selectivity was tolerant to distance, implying tuning for relative disparity gradients or 3D pose. However, selectivity for stereoscopically defined 3D curved (convex versus concave) surfaces was not tolerant to distance. Most recently, V4 was shown to contain 3D volumetric shape representations signaled by shading, reflectance, and disparity (Srinath et al. 2021). These results together suggest that V4 contains heterogeneous 3D representations, not unlike areas within the dorsal pathway.

Ventral visual processing continues after V4 in posterior inferotemporal (IT) cortex, also called TEO, which additionally receives input from V3A (Distler et al. 1993). Within TEO, 2D contour

representations like those found in V4 are combined to achieve more complex shape selectivity (Brincat & Connor 2004). However, little is known about TEO's 3D selectivity. Neuroimaging work has linked disparity-sensitive domains in TEO to scene processing and identified a region that responded more strongly to stereoscopically defined curved than flat surfaces (Kornblith et al. 2013, Van Dromme et al. 2016, Verhoef et al. 2015). However, electrophysiological recordings from that region found few neurons selective for stereoscopically defined surfaces (Alizadeh et al. 2018b). Such processing may occur in other parts of TEO, as perhaps suggested by a study that observed reduced stereoacuity after TEO was bilaterally lesioned (Cowey & Porter 1979). It is also possible that 3D processing within TEO relies predominantly on perspective and/or other monocular cues. Unfortunately, the current dearth of relevant studies precludes clear conclusions about TEO's role in 3D vision.

Anterior IT cortex, also called TE, near the end stages of the ventral pathway, receives input from V4 and TEO (Distler et al. 1993, Ungerleider et al. 2008) and is connected with parietal, prefrontal, and temporal areas (Saleem et al. 2000, Webster et al. 1994). The well-known modular organization of TE includes two 3D selective domains: (*a*) the ventral bank of the superior temporal sulcus (STSv or TEs) and (*b*) the lateral aspect of TE (TEd). These domains have been associated with 3D object (Janssen et al. 2000a,b; Yamane et al. 2008) and scene (Vaziri & Connor 2016, Vaziri et al. 2014) processing, respectively.

Neurons in STSv show selectivity for curved and flat 3D surfaces (Janssen et al. 1999, 2000a; Liu et al. 2004), as well as more complex shapes (Sereno et al. 2002, Tanaka et al. 2001, Vaziri et al. 2014) signaled by stereoscopic, perspective, and shading cues. Notably, the correspondence problem has largely been solved at the level of STSv (Janssen et al. 2003), and the neurons are less responsive to stimuli lacking stereoscopic and shading cues, suggesting that they may integrate multiple cues to achieve more precise 3D representations. Shape selectivity in STSv also shows tolerance to translations and rotations in the image plane (Yamane et al. 2008), as well as to distance (Janssen et al. 2000b), implying high-level feature selectivity. A role for STSv in 3D perception is further supported by functional correlations and causal relationships between neuronal activity in the area and behavioral reports during stereoscopic 3D curvature discrimination (Verhoef et al. 2010, 2012).

In contrast to STSv, neurons in TEd show selectivity for large landscape-like surfaces signaled by texture and shading cues rather than stereoscopic cues (Vaziri & Connor 2016, Vaziri et al. 2014). For these neurons, preferred stimuli include large surfaces consistent with floors, walls, or ceilings. It is conceivable that this selectivity arises through hierarchical processing of earlier scene representations in TEO (Kornblith et al. 2013, Verhoef et al. 2015) (**Figure 5**, *bottom right*). The interpretation that TEd neurons are selective for 3D scene structure is bolstered by an overrepresentation of horizontal (floors, ceilings) and vertical (walls) orientations, which could provide an environmental frame signal that supports balance control (Dakin & Rosenberg 2018). It would therefore be interesting to test if vestibular/proprioceptive signals are utilized at the level of TEd to achieve gravity-aligned scene representations, similar to the gravity-aligned object representations in CIP (Rosenberg & Angelaki 2014a).

In future work, it will be important to explore whether there are functional interactions between STSv and TEd during behavioral tasks. For example, such interactions may facilitate the recognition of objects presented in a larger scene context (Oliva & Torralba 2007). Interactions between these areas may also enable mnemonic associations between objects and the environment, which could support the formation and updating of cognitive maps based on the locations of key resources (Connor & Knierim 2017).

CONCLUSIONS, OPEN QUESTIONS, AND FUTURE DIRECTIONS

Three-dimensional visual information is processed in parallel and hierarchically within the dorsal and ventral pathways. At a finer scale, particular 3D features are processed by specialized channels within these pathways. While functional distinctions between these 3D-fugal processing channels may be drawn, they also have common origins, they share processing nodes, and substantial communication likely occurs between them (**Figure 5**). The significance of this overlap and information sharing remains surprisingly unexplored.

A Connection Between the Provenance of Dorsal and Ventral Three-Dimensional Representations and Human Spatial Reasoning?

A textbook view of the primate visual system is that the dorsal and ventral pathways are rooted in distinct processing domains within V2. Downstream of V2, area V3A has been associated with the dorsal processing of 3D object pose (Doudlah et al. 2022, Nakamura et al. 2001) and, recently, 3D motion (Nakhla et al. 2021). However, V3A also projects to ventral areas V4 and TEO (Distler et al. 1993, Ungerleider et al. 2008). Based on the response properties of V3A and its downstream targets, it is plausible that V3A is a key branchpoint from which multiple (dorsal and ventral) 3D-fugal processing channels originate (**Figure 5**). Comparative analyses further suggest that the human homolog of V3A expanded disproportionately compared to neighboring areas (Van Essen et al. 2001), and human neuroimaging data strongly implicate V3A in 3D vision (Henderson et al. 2019). Given these observations, it is intriguing to speculate that V3A serves a central role in primate 3D vision, and that its expansion may have supported the evolution of advanced 3D spatial reasoning skills.

Standardizing Experimental Paradigms

Explicating how retinal images are transformed into robust 3D representations has been challenging, in part due to the tendency for different studies to use different stimuli, tasks, and analytical techniques. Although they are currently somewhat rare, direct cross-area comparisons have proven fruitful. In the dorsal pathway, this has revealed increased 3D pose selectivity between V3A and CIP (Doudlah et al. 2022, Elmore et al. 2019) and a pronounced transformation from 2D to 3D motion selectivity between MT and FST (Thompson et al. 2022). Within the ventral pathway, this strategy demonstrated a stark difference in the processing of 3D objects versus scenes in STSv and TEd (Vaziri et al. 2014). Direct comparisons have also revealed key differences across the visual pathways. For instance, disparity selectivity was found to be stronger and faster in MT than in V4, whereas the rejection of false matches was more complete in V4 (Yoshioka et al. 2021). Such examples highlight how future work could benefit from using standardized experimental designs to investigate the functional specializations and computational differences between brain areas supporting 3D vision.

Functional Implications of Cross-Area Communication for Three-Dimensional Vision

Perception of the 3D world is remarkably stable despite the sensory consequences of eye movements, self-motion, and the volatility of individual objects. This stability suggests that various processing channels supporting 3D perception are inextricably linked. Indeed, functional, anatomical, and effective connectivity data suggest that pronounced 3D information sharing occurs (Janssen et al. 2018). For example, reversible inactivation of CIP was found to reduce STSv activity, although electrical microstimulation of CIP did not drive STSv (Van Dromme et al. 2016).

This may indicate that CIP influences 3D processing in STSv via connections with AIP (Borra et al. 2008, Webster et al. 1994) (**Figure 5**). Notably, 3D object shape is processed in both AIP and STSv, but whereas this selectivity emerges more quickly in AIP, it is finer in STSv (Srivastava et al. 2009, Verhoef et al. 2010). This distinction may reflect differences in the temporal urgency of action versus the refinement of 3D object representations. At longer timescales, interactions between STSv and AIP may support 3D spatial reasoning processes in which object identity and prehensile signals together define object affordances. As another example, connections between FST and 3D selective areas in the ventral pathway (Boussaoud et al. 1990, Distler et al. 1993, Saleem et al. 2000, Ungerleider et al. 2008) (**Figure 5**) may support object selectivity in FST (Bogadhi et al. 2021) and allow for the binding of object identity and 3D motion relative to specific environmental features.

Beyond characterizing area-specific contributions to 3D vision, simultaneous recordings from interconnected areas (especially during 3D perceptual and sensorimotor tasks) promise to illuminate the functional consequences of cross-area communication. Along this line, one exciting research direction to pursue is clarifying how and what 3D information is shared between the dorsal and ventral pathways. Intriguingly, recent work has suggested that dorsal–ventral communication is asymmetric, such that dorsal object representations may have a larger influence on ventral representations than vice versa (Ayzenberg & Behrmann 2022, Premereur & Janssen 2020). One possibility generally consistent with the existing literature is that the dorsal pathway quickly computes high-level but comparatively coarse 3D object representations to support action and then relays this information to the ventral pathway to contribute to the computation of more refined 3D representations that include detailed features.

Lastly, given the success of geometric approaches to defining hypothesis-driven tests of 3D pose tuning (Chang et al. 2020a, Doudlah et al. 2022, Rosenberg et al. 2013), future efforts to explicate the parallel and hierarchical transformations that support 3D vision would likely benefit from a unified framework that parameterizes both 3D shape and pose. Indeed, a synergistic combination of theoretically driven experimental paradigms, simultaneous recordings from multiple interconnected areas, and high-level 3D perceptual and sensorimotor tasks would provide a powerful approach to further elucidate the neural computations that enable primates to uniquely thrive in the 3D world.

DISCLOSURE STATEMENT

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

ACKNOWLEDGMENTS

We thank Peter Janssen and Bas Rokers for helpful comments on this review. A.R. was supported by the Greater Milwaukee Foundation (Shaw Scientist Award) and National Institutes of Health (EY029438, EY035005, NS128586). L.W.T. was supported by the McPherson Eye Research Institute (Graduate Student Support Initiative) and National Institutes of Health (NS105602). R.D. was supported by the National Institutes of Health (EY027721). Further support was provided by National Institutes of Health grant P51OD011106 to the Wisconsin National Primate Research Center.

LITERATURE CITED

Albright TD, Desimone R, Gross CG. 1984. Columnar organization of directionally selective cells in visual area MT of the macaque. *J. Neurophysiol.* 51:16–31

Alizadeh AM, Van Dromme I, Verhoef BE, Janssen P. 2018a. Caudal Intraparietal Sulcus and three-dimensional vision: a combined functional magnetic resonance imaging and single-cell study. *NeuroImage* 166:46–59

Alizadeh AM, Van Dromme IC, Janssen P. 2018b. Single-cell responses to three-dimensional structure in a functionally defined patch in macaque area TEO. *J. Neurophysiol.* 120:2806–18

Angelaki DE, Gu Y, Deangelis GC. 2011. Visual and vestibular cue integration for heading perception in extrastriate visual cortex. *J. Physiol.* 589:825–33

Anzai A, Chowdhury SA, DeAngelis GC. 2011. Coding of stereoscopic depth information in visual areas V3 and V3A. *J. Neurosci.* 31:10270–82

Ayzenberg V, Behrmann M. 2022. Does the brain's ventral visual pathway compute object shape? *Trends Cogn. Sci.* 26:1119–32

Banker SM, Ramphal B, Pagliaccio D, Thomas L, Rosen E, et al. 2020. Spatial network connectivity and spatial reasoning ability in children with nonverbal learning disability. *Sci. Rep.* 10:561

Barton RA. 2004. From the cover: binocularity and brain evolution in primates. *PNAS* 101:10113–15

Bogadhi AR, Katz LN, Bollimunta A, Leopold DA, Krauzlis RJ. 2021. Midbrain activity shapes high-level visual properties in the primate temporal cortex. *Neuron* 109:690–99.e5

Bonnen K, Czuba TB, Whritner JA, Kohn A, Huk AC, Cormack LK. 2020. Binocular viewing geometry shapes the neural representation of the dynamic three-dimensional environment. *Nat. Neurosci.* 23:113–21

Born RT, Bradley DC. 2005. Structure and function of visual area MT. *Annu. Rev. Neurosci.* 28:157–89

Borra E, Belmalih A, Calzavara R, Gerbella M, Murata A, et al. 2008. Cortical connections of the macaque anterior intraparietal (AIP) area. *Cereb. Cortex* 18:1094–111

Boussaoud D, Ungerleider LG, Desimone R. 1990. Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *J. Comp. Neurol.* 296:462–95

Brincat SL, Connor CE. 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7:880–86

Brooks KR. 2002. Interocular velocity difference contributes to stereomotion speed perception. *J. Vis.* 2:218–31

Chang TY, Doudlah R, Kim B, Sunkara A, Thompson LW, et al. 2020a. Functional links between sensory representations, choice activity, and sensorimotor associations in parietal cortex. *eLife* 9:e57968

Chang TY, Thompson LW, Doudlah R, Kim B, Sunkara A, Rosenberg A. 2020b. Optimized but not maximized cue integration for 3D visual perception. *eNeuro* 7:ENEURO.0411-19.2019

Clark JJ, Yuille AL. 1990. *Data Fusion for Sensory Information Processing*. Springer Int. Ser. Eng. Comput. Sci. 105. Berlin: Springer

Connor CE, Knierim JJ. 2017. Integration of objects and space in perception and memory. *Nat. Neurosci.* 20:1493–503

Cooper EA, van Ginkel M, Rokers B. 2016. Sensitivity and bias in the discrimination of two-dimensional and three-dimensional motion direction. *J. Vis.* 16:5

Cormack LK, Czuba TB, Knöll J, Huk AC. 2017. Binocular mechanisms of 3D motion processing. *Annu. Rev. Vis. Sci.* 3:297–318

Cowan NJ, Chang DE. 2005. Geometric visual servoing. *IEEE Trans. Robot.* 21:1128–38

Cowan NJ, Weingarten JD, Koditschek DE. 2002. Visual servoing via navigation functions. *IEEE Trans. Robot. Autom.* 18:521–33

Cowey A, Porter J. 1979. Brain damage and global stereopsis. *Proc. R. Soc. Lond. B* 204:399–407

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

Cumming BG, Parker AJ. 1994. Binocular mechanisms for detecting motion-in-depth. *Vis. Res.* 34:483–95

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

Czuba TB, Huk AC, Cormack LK, Kohn A. 2014. Area MT encodes three-dimensional motion. *J. Neurosci.* 34:15522–33

Czuba TB, Rokers B, Guillet K, Huk AC, Cormack LK. 2011. Three-dimensional motion aftereffects reveal distinct direction-selective mechanisms for binocular processing of motion through depth. *J. Vis.* 11:18

Dakin CJ, Rosenberg A. 2018. Gravity estimation and verticality perception. *Handb. Clin. Neurol.* 159:43–59
De Valois RL, Yund EW, Hepler N. 1982. The orientation and direction selectivity of cells in macaque visual cortex. *Vis. Res.* 22:531–44
DeAngelis GC, Uka T. 2003. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J. Neurophysiol.* 89:1094–111
Desimone R, Albright TD, Gross CG, Bruce C. 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* 4:2051–62
Distler C, Boussaoud D, Desimone R, Ungerleider LG. 1993. Cortical connections of inferior temporal area TEO in macaque monkeys. *J. Comp. Neurol.* 334:125–50
Doudlah R, Chang TY, Thompson LW, Kim B, Sunkara A, Rosenberg A. 2022. Parallel processing, hierarchical transformations, and sensorimotor associations along the 'where' pathway. *eLife* 11:e78712
Du Q, Ts'o D. 2013. *Responses in area V2 to disparity gradients: a role in 3D surface analysis*. Presented at the Society for Neuroscience 2013, Nov. 9–13, San Diego, CA
Durand JB, Nelissen K, Joly O, Wardak C, Todd JT, et al. 2007. Anterior regions of monkey parietal cortex process visual 3D shape. *Neuron* 55:493–505
Eifuku S, Wurtz RH. 1998. Response to motion in extrastriate area MSTl: center-surround interactions. *J. Neurophysiol.* 80:282–96
Elmore LC, Rosenberg A, DeAngelis GC, Angelaki DE. 2019. Choice-related activity during visual slant discrimination in macaque CIP but not V3A. *eNeuro* 6:ENEURO.0248-18.2019
Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1:1–47
Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. 2011. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat. Neurosci.* 15:146–54
Fulvio JM, Ji M, Thompson LW, Rosenberg A, Rokers B. 2020. Cue-dependent effects of VR experience on motion-in-depth sensitivity. *PLOS ONE* 15:e0229929
Galletti C, Battaglini PP. 1989. Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *J. Neurosci.* 9:1112–25
Galletti C, Battaglini PP, Fattori P. 1990. "Real-motion" cells in area V3A of macaque visual cortex. *Exp. Brain Res.* 82:67–76
Gaska JP, Jacobson LD, Pollen DA. 1987. Response suppression by extending sine-wave gratings within the receptive fields of neurons in visual cortical area V3A of the macaque monkey. *Vis. Res.* 27:1687–92
Gaska JP, Jacobson LD, Pollen DA. 1988. Spatial and temporal frequency selectivity of neurons in visual cortical area V3A of the macaque monkey. *Vis. Res.* 28:1179–91
Gibson JJ. 1947. *Motion Picture Testing and Research*. Washington, DC: U.S. Gov. Print. Off.
Greenwald HS, Knill DC. 2009. Orientation disparity: a cue for 3D orientation? *Neural Comput.* 21:2581–604
Harris JM, Wilcox LM. 2009. The role of monocularly visible regions in depth and surface perception. *Vis. Res.* 49:2666–85
Hegdé J, Van Essen DC. 2005. Role of primate visual area V4 in the processing of 3-D shape characteristics defined by disparity. *J. Neurophysiol.* 94:2856–66
Hegdé J, Van Essen DC. 2007. A comparative study of shape representation in macaque visual areas v2 and v4. *Cereb. Cortex* 17:1100–16
Héjja-Brichard Y, Rima S, Rapha E, Durand JB, Cottereau BR. 2020. Stereomotion processing in the nonhuman primate brain. *Cereb. Cortex* 30:4528–43
Held RT, Cooper EA, O'Brien JF, Banks MS. 2010. Using blur to affect perceived distance and size. *ACM Trans. Graph.* 29:19
Henderson M, Vo V, Chunharas C, Sprague T, Serences J. 2019. Multivariate analysis of BOLD activation patterns recovers graded depth representations in human visual and parietal cortex. *eNeuro* 6:ENEURO.0362-18.2019
Hillis JM, Watt SJ, Landy MS, Banks MS. 2004. Slant from texture and disparity cues: optimal cue combination. *J. Vis.* 4:967–92
Hinkle DA, Connor CE. 2001. Disparity tuning in macaque area V4. *Neuroreport* 12:365–69
Hinkle DA, Connor CE. 2002. Three-dimensional orientation tuning in macaque area V4. *Nat. Neurosci.* 5:665–70

Howard IP, Rogers BJ. 1995. *Binocular Vision and Stereopsis*. Oxford, UK: Oxford Univ. Press
Humphrey GK, Khan SC. 1992. Recognizing novel views of three-dimensional objects. *Can. J. Psychol.* 46:170–90
Ilg UJ, Schumann S, Thier P. 2004. Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron* 43:145–51
Janssen P, Verhoef BE, Premereur E. 2018. Functional interactions between the macaque dorsal and ventral visual pathways during three-dimensional object vision. *Cortex* 98:218–27
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* 37:693–701
Janssen P, Vogels R, Orban GA. 1999. Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *PNAS* 96:8217–22
Janssen P, Vogels R, Orban GA. 2000a. Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* 288:2054–56
Janssen P, Vogels R, Orban GA. 2000b. Three-dimensional shape coding in inferior temporal cortex. *Neuron* 27:385–97
Kim HR, Angelaki DE, DeAngelis GC. 2016. The neural basis of depth perception from motion parallax. *Philos. Trans. R. Soc. Lond. B* 371:0256
Kim HR, Angelaki DE, DeAngelis GC. 2022. A neural mechanism for detecting object motion during self-motion. *eLife* 11:e74971
Knill DC. 1998. Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vis. Res.* 38:1655–82
Knill DC, Richards W. 1996. *Perception as Bayesian Inference*. Cambridge, UK: Cambridge Univ. Press
Knill DC, Saunders JA. 2003. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vis. Res.* 43:2539–58
Kobatake E, Tanaka K. 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* 71:856–67
Koenderink JJ, van Doorn AJ. 2004. Shape and shading. In *The Visual Neurosciences*, ed. LM Chalupa, JS Werner, pp. 1090–105. Cambridge, MA: MIT Press
Kornblith S, Cheng X, Ohayon S, Tsao DY. 2013. A network for scene processing in the macaque temporal lobe. *Neuron* 79:766–81
Kotake Y, Morimoto H, Okazaki Y, Fujita I, Tamura H. 2009. Organization of color-selective neurons in macaque visual area V4. *J. Neurophysiol.* 102:15–27
Lagae L, Maes H, Raiguel S, Xiao DK, Orban GA. 1994. Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. *J. Neurophysiol.* 71:1597–626
Liu Y, Vogels R, Orban GA. 2004. Convergence of depth from texture and depth from disparity in macaque inferior temporal cortex. *J. Neurosci.* 24:3795–800
Longuet-Higgins HC, Prazdny K. 1980. The interpretation of a moving retinal image. *Proc. R. Soc. Lond. B* 208:385–97
Ma WJ, Beck JM, Latham PE, Pouget A. 2006. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9:1432–38
Mather G. 1996. Image blur as a pictorial depth cue. *Proc. R. Soc. Lond. B* 263:169–72
Maunsell JH, Van Essen DC. 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* 49:1148–67
Mishkin M, Ungerleider LG, Macko KA. 1983. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6:414–17
Movshon JA, Adelson EH, Gizzi MS, Newsome WT. 1985. The analysis of visual moving patterns. In *Pattern Recognition Mechanisms*, ed. C Chagas, R Gattass, C Gross, pp. 117–51. Berlin: Springer
Movshon JA, Newsome WT. 1996. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J. Neurosci.* 16:7733–41
Murata A, Gallese V, Luppino G, Kaseda M, Sakata H. 2000. Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *J. Neurophysiol.* 83:2580–601

Mysore SG, Vogels R, Raiguel SE, Todd JT, Orban GA. 2010. The selectivity of neurons in the macaque fundus of the superior temporal area for three-dimensional structure from motion. *J. Neurosci.* 30:15491–508
Nadler JW, Angelaki DE, DeAngelis GC. 2008. A neural representation of depth from motion parallax in macaque visual cortex. *Nature* 452:642–45
Nakamura H, Kuroda T, Wakita M, Kusunoki M, Kato A, et al. 2001. From three-dimensional space vision to prehensile hand movements: the lateral intraparietal area links the area V3A and the anterior intraparietal area in macaques. *J. Neurosci.* 21:8174–87
Nakamura K, Colby CL. 2000. Visual, saccade-related, and cognitive activation of single neurons in monkey extrastriate area V3A. *J. Neurophysiol.* 84:677–92
Nakamura K, Colby CL. 2002. Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *PNAS* 99:4026–31
Nakhla N, Korkian Y, Krause MR, Pack CC. 2021. Neural selectivity for visual motion in macaque area V3A. *eNeuro* 8:ENEURO.0383-20.2020
Nguyenkim JD, DeAngelis GC. 2003. Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J. Neurosci.* 23:7117–28
Nover H, Anderson CH, DeAngelis GC. 2005. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J. Neurosci.* 25:10049–60
Oliva A, Torralba A. 2007. The role of context in object recognition. *Trends Cogn. Sci.* 11:520–27
Orban GA, Kennedy H, Bullier J. 1986. Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *J. Neurophysiol.* 56:462–80
Oruç I, Maloney LT, Landy MS. 2003. Weighted linear cue combination with possibly correlated error. *Vis. Res.* 43:2451–68
Parker AJ. 2007. Binocular depth perception and the cerebral cortex. *Nat. Rev. Neurosci.* 8:379–91
Pasternak T, Tadin D. 2020. Linking neuronal direction selectivity to perceptual decisions about visual motion. *Annu. Rev. Vis. Sci.* 6:335–62
Pasupathy A, Connor CE. 2001. Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* 86:2505–19
Poggio GF, Talbot WH. 1981. Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J. Physiol.* 315:469–92
Pohl W. 1973. Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys. *J. Comp. Physiol. Psychol.* 82:227–39
Premereur E, Janssen P. 2020. Effective connectivity reveals an interconnected inferotemporal network for three-dimensional structure processing. *J. Neurosci.* 40:8501–12
Qamar AT, Cotton RJ, George RG, Beck JM, Prezhdo E, et al. 2013. Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *PNAS* 110:20332–37
Rogers BJ. 1993. Motion parallax and other dynamic cues for depth in humans. *Rev. Oculomot. Res.* 5:119–37
Rokers B, Cormack LK, Huk AC. 2008. Strong percepts of motion through depth without strong percepts of position in depth. *J. Vis.* 8:6
Rosenberg A, Angelaki DE. 2014a. Gravity influences the visual representation of object tilt in parietal cortex. *J. Neurosci.* 34:14170–80
Rosenberg A, Angelaki DE. 2014b. Reliability-dependent contributions of visual orientation cues in parietal cortex. *PNAS* 111:18043–48
Rosenberg A, Cowan NJ, Angelaki DE. 2013. The visual representation of 3D object orientation in parietal cortex. *J. Neurosci.* 33:19352–61
Rosenberg A, Patterson JS, Angelaki DE. 2015. A computational perspective on autism. *PNAS* 112:9158–65
Rosenberg A, Wallisch P, Bradley DC. 2008. Responses to direction and transparent motion stimuli in area FST of the macaque. *Vis. Neurosci.* 25:187–95
Sabatini SP, Solari F. 2004. Emergence of motion-in-depth selectivity in the visual cortex through linear combination of binocular energy complex cells with different ocular dominance. *Neurocomputing* 58–60:865–72
Saleem KS, Suzuki W, Tanaka K, Hashikawa T. 2000. Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. *J. Neurosci.* 20:5083–101

Sanada TM, DeAngelis GC. 2014. Neural representation of motion-in-depth in area MT. *J. Neurosci.* 34:15508–21
Sasaki R, Anzai A, Angelaki DE, DeAngelis GC. 2020. Flexible coding of object motion in multiple reference frames by parietal cortex neurons. *Nat. Neurosci.* 23:1004–15
Sauvan XM, Peterhans E. 1999. Orientation constancy in neurons of monkey visual cortex. *Vis. Cogn.* 6:43–54
Seilheimer RL, Rosenberg A, Angelaki DE. 2014. Models and processes of multisensory cue combination. *Curr. Opin. Neurobiol.* 25:38–46
Sereno ME, Trinath T, Augath M, Logothetis NK. 2002. Three-dimensional shape representation in monkey cortex. *Neuron* 33:635–52
Shimojo S, Nakayama K. 1990. Real world occlusion constraints and binocular rivalry. *Vis. Res.* 30:69–80
Srinath R, Emonds A, Wang Q, Lempel AA, Dunn-Weiss E, et al. 2021. Early emergence of solid shape coding in natural and deep network vision. *Curr. Biol.* 31:51–65.e5
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.* 29:10613–26
Stevens KA. 1981. The information content of texture gradients. *Biol. Cybern.* 42:95–105
Sunkara A, DeAngelis GC, Angelaki DE. 2015. Role of visual and non-visual cues in constructing a rotation-invariant representation of heading in parietal cortex. *eLife* 4:e04693
Sunkara A, DeAngelis GC, Angelaki DE. 2016. Joint representation of translational and rotational components of optic flow in parietal cortex. *PNAS* 113:5077–82
Tanabe S, Umeda K, Fujita I. 2004. Rejection of false matches for binocular correspondence in macaque visual cortical area V4. *J. Neurosci.* 24:8170–80
Tanaka H, Uka T, Yoshiyama K, Kato M, Fujita I. 2001. Processing of shape defined by disparity in monkey inferior temporal cortex. *J. Neurophysiol.* 85:735–44
Tanaka K, Sugita Y, Moriya M, Saito H. 1993. Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *J. Neurophysiol.* 69:128–42
Theys T, Pani P, van Loon J, Goffin J, Janssen P. 2012. Selectivity for three-dimensional shape and grasping-related activity in the macaque ventral premotor cortex. *J. Neurosci.* 32:12038–50
Thomas OM, Cumming BG, Parker AJ. 2002. A specialization for relative disparity in V2. *Nat. Neurosci.* 5:472–78
Thompson LW, Ji M, Rokers B, Rosenberg A. 2019. Contributions of binocular and monocular cues to motion-in-depth perception. *J. Vis.* 19:2
Thompson LW, Kim B, Rokers B, Rosenberg A. 2022. *Hierarchical processing of 3D motion in macaque MT and FST*. Presented at the Society for Neuroscience 2022, Nov. 12–16, San Diego, CA
Thompson LW, Kim B, Zhu Z, Rokers B, Rosenberg A. 2021. Perspective cues make eye-specific contributions to 3D motion perception. *J. Cogn. Neurosci.* 34:192–208
Todd JT, Mingolla E. 1983. Perception of surface curvature and direction of illumination from patterns of shading. *J. Exp. Psychol. Hum. Percept. Perform.* 9:583–95
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* 39:555–68
Tsutsui K, Jiang M, Sakata H, Taira M. 2003. Short-term memory and perceptual decision for three-dimensional visual features in the caudal intraparietal sulcus (Area CIP). *J. Neurosci.* 23:5486–95
Tsutsui K, Jiang M, Yara K, Sakata H, Taira M. 2001. Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. *J. Neurophysiol.* 86:2856–67
Tsutsui K, Sakata H, Naganuma T, Taira M. 2002. Neural correlates for perception of 3D surface orientation from texture gradient. *Science* 298:409–12
Umeda K, Tanabe S, Fujita I. 2007. Representation of stereoscopic depth based on relative disparity in macaque area V4. *J. Neurophysiol.* 98:241–52
Ungerleider LG, Desimone R. 1986. Cortical connections of visual area MT in the macaque. *J. Comp. Neurol.* 248:190–222
Ungerleider LG, Galkin TW, Desimone R, Gattass R. 2008. Cortical connections of area V4 in the macaque. *Cereb. Cortex* 18:477–99
Uttal DH, Cohen CA. 2012. Spatial thinking and STEM education: when, why, and how? In *The Psychology of Learning and Motivation*, ed. B Ross, pp. 147–81. Amsterdam: Elsevier

Van Dromme IC, Premereur E, Verhoef BE, Vanduffel W, Janssen P. 2016. Posterior parietal cortex drives inferotemporal activations during three-dimensional object vision. *PLOS Biol.* 14:e1002445

Van Essen DC, Lewis JW, Drury HA, Hadjikhani N, Tootell RBH, et al. 2001. Mapping visual cortex in monkeys and humans using surface-based atlases. *Vis. Res.* 41:1359–78

Vaziri S, Carlson ET, Wang Z, Connor CE. 2014. A channel for 3D environmental shape in anterior inferotemporal cortex. *Neuron* 84:55–62

Vaziri S, Connor CE. 2016. Representation of gravity-aligned scene structure in ventral pathway visual cortex. *Curr. Biol.* 26:766–74

Verhoef BE, Bohon KS, Conway BR. 2015. Functional architecture for disparity in macaque inferior temporal cortex and its relationship to the architecture for faces, color, scenes, and visual field. *J. Neurosci.* 35:6952–68

Verhoef BE, Vogels R, Janssen P. 2010. Contribution of inferior temporal and posterior parietal activity to three-dimensional shape perception. *Curr. Biol.* 20:909–13

Verhoef BE, Vogels R, Janssen P. 2012. Inferotemporal cortex subserves three-dimensional structure categorization. *Neuron* 73:171–82

Verhoef BE, Vogels R, Janssen P. 2016. Binocular depth processing in the ventral visual pathway. *Philos. Trans. R. Soc. Lond. B* 371:20150259

Webster MJ, Bachevalier J, Ungerleider LG. 1994. Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb. Cortex* 4:470–83

Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE. 2008. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11:1352–60

Yoshioka TW, Doi T, Abdolrahmani M, Fujita I. 2021. Specialized contributions of mid-tier stages of dorsal and ventral pathways to stereoscopic processing in macaque. *eLife* 10:e58749

Zeki SM. 1974. Cells responding to changing image size and disparity in the cortex of the rhesus monkey. *J. Physiol.* 242:827–41