Published in final edited form as:

Annu Rev Vis Sci. 2017 September 15; 3: 297–318. doi:10.1146/annurev-vision-102016-061259.

Binocular Mechanisms of 3D Motion Processing

Lawrence K. Cormack, Thaddeus B. Czuba, Jonas Knöll, and Alexander C. Huk

The University of Texas at Austin, Texas 78757

Abstract

The visual system must recover important properties of the external environment if its host is to survive. Because the retinae are effectively two-dimensional but the world is three-dimensional (3D), the patterns of stimulation both within and across the eyes must be used to infer the distal stimulus—the environment—in all three dimensions. Moreover, animals and elements in the environment move, which means the input contains rich temporal information. Here, in addition to reviewing the literature, we discuss how and why prior work has focused on purported isolated systems (e.g., stereopsis) or cues (e.g., horizontal disparity) that do not necessarily map elegantly on to the computations and complex patterns of stimulation that arise when visual systems operate within the real world. We thus also introduce the binoptic flow field (BFF) as a description of the 3D motion information available in realistic environments, which can foster the use of ecologically valid yet well-controlled stimuli. Further, it can help clarify how future studies can more directly focus on the computations and stimulus properties the visual system might use to support perception and behavior in a dynamic 3D world.

Keywords

motion; depth; disparity; optic flow; binocular; geometry

1. INTRODUCTION

Animals, by definition, move. When an animal moves, this creates both relative motion of the environment with respect to the animal and relative motion of the animal with respect to all other animals and objects in the vicinity. Given the availability of light and a transparent medium, vision is a great way to detect such motion. In fact, if a terrestrial organism has light in its normal environment but does not respond to visual motion in some way, then it is a safe bet that it is not an animal. Interestingly, all animals with eyes do seem to share a common basic behavior, which is a defensive response to sudden visual motion (Land & Nilsson 2012); in many cases (e.g., most bivalves), this is the only visuomotor response the animal possesses. Judging from such rudimentary but extant visual systems, it would seem that before any mechanisms existed to compute three-dimensional (3D) motion explicitly, all motion was assumed to be threatening—or, in other words, 3D motion toward the animal.

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

Thus, to paraphrase Nilsson (2006), early eyes were probably burglar alarms long before they were security cameras.

We thus find it likely that the utility of motion detection, however primitive, provided the initial selection pressure for the formation of eyes and visual systems. Even today, one can conjecture that the most important function of the primate visual system, in terms of individual survival, is extracting information regarding the relative position and movement of objects through the three-dimensional (3D) environment. The ubiquitous combination of object and observer motions imply that the visual scene as encountered by the two eyes is utterly dynamic, and thus appropriate actions and navigation will depend critically on such 3D motion processing that likely occurs in the dorsal stream of the primate visual system. There is no doubt that the ventral stream of visual processing—responsible for functions such as object and face identification—is important to modern primates, but we can nonetheless dodge suddenly threatening objects without bothering to recognize them.

Despite the arguable primacy of 3D motion as a visual feature, the perception and neural processing of motion and depth have generally been separated and treated independently as simplified model systems for understanding how the nervous system processes information. Although such studies have generated many insights into the neural basis of perception, their modular successes do not provide a unified understanding of how motion and depth information are combined and used to support behaviors that need not cleanly distinguish between these types of information.

In this review, we attempt to recombine motion and depth in a tractable and fruitful framework for future work. First, we briefly explain the historical circumstances that led to the over-abstraction of 3D motion processing. Second, we explicate the richness of motion and depth information available to the visual system. In doing so, we introduce the concept of the binoptic flow field (BFF), which is a comprehensive geometric model that describes the information available to a binocular observer given any real-world scene structure and 3D movement. We then selectively review both psychophysical and neurophysiological literature through the lens of the BFF. This exercise allows us to interpret existing results and suggest future work that focuses more fundamentally on how the visual system encodes and decodes environmental information, thus expanding the current emphasis on how it processes differential information presented to the two eyes.

1.1. Early Approaches to Motion and Depth

One of the most elegant insights into visual processing followed from the random-element stereogram (Julesz 1971). In these visual versions of white noise, neither eye's view contains any identifiable structure. However, by introducing horizontal offsets between the random elements in some locations, the stimulus yields a percept of depth or depth structure when viewed stereoscopically. Importantly, these stimuli are remarkable because they demonstrate that disparity processing does not require monocularly identifiable elements or recognizable objects upon which to compute disparities; instead, the disparity extraction appears to be a low-level cross-correlation-like function, for which dense, random texture is, perhaps counterintuitively, actually the best input (Nishihara 1984). This implies an early extraction of disparities, likely at the point of initial binocular combination in primary visual cortex—

as opposed to, say, after line segments are reconstructed in larger receptive fields or after a scene is parsed into meaningful elements. It is at this early site of computing disparities that the visual system becomes effectively “cyclopean.” Understandably, the result of this was the unspoken assumption that, because stereopsis *could* operate using only texture primitives, then it most likely *only* operated using texture primitives. This, in turn, helped cement the notion that stereopsis was a module unto itself, operating independently of other aspects of vision, such as motion.

Likewise, the domain of motion has remained fairly separate from that of stereopsis and has grown into its own model system. This was in part driven, as was the case for stereopsis, by the introduction of a particularly elegant stimulus—in this case, the random-dot kinetogram (Anstis 1970), which is effectively one of Julesz’s stereograms presented to one eye over time instead of binocularly. Later, sparser dot fields were introduced as stimuli (Braddick 1974, Morgan & Ward 1980), and it was subsequently realized that sparse, random-dot displays had an interesting advantage over other motion stimuli. The power of these displays, as opposed to discrete stimuli (like moving bars) or periodic stimuli (such as drifting gratings), is that they allow for both (*a*) the direct titration of motion strength by varying the fraction of dots that move consistently, versus those that move along random trajectories, and (*b*) control over trackable dot duration and/or the minimization of trackable elements by varying the lifetime of individual, coherently moving dots (Newsome & Paré 1988, Williams & Sekuler 1984).

The use of random-dot kinetograms facilitated a series of experiments that linked the perception of frontoparallel (2D) direction with the neural signals in extrastriate visual areas MT (middle temporal area) and MST (medial superior temporal area). Although it was already known that MT contained a notably large fraction of direction-selective neurons (Albright 1984, Zeki 1974b), these experiments both quantitatively and causally linked the sensitivity of these neurons to varying degrees of motion strength to performance on a direction-discrimination task (Britten et al. 1992, 1993; Newsome & Paré 1988; Shadlen & Newsome 2001). Disparity tuning was also found in MT, but influential early work on 3D motion concluded that responses in MT were wholly explained by a combination of 2D motion tuning and static disparity tuning (Maunsell & Van Essen 1983). This led to a zeitgeist in which motion processing was thought to be inherently 2D but was processed within various disparity-defined depth planes. Such a computational strategy would be eerily convenient, given the perfect mapping onto the flat, frontoparallel displays generally used to probe the visual system.

This rough segregation of the stereo and motion domains has even been reflected in the approaches taken to the study of 3D motion itself. Initial studies took the random-element stereogram and made the disparities dynamic and, by virtue of updating the random elements at each frame, were able to maintain the purely cyclopean (disparity-based) nature of the stimulus. Because subjects can perceive 3D motion in such stimuli (Julesz 1971), the notion that changing disparities (CDs) (in the absence of any coherent retinal motions) were the primary stimulus for motion through depth was affirmed. Following this, the CD cue was widely and often implicitly assumed to be the only binocular cue for motion through depth.

Such early studies of disparity-based depth motion were then complemented by work that incorporated monocular velocities. Just as an object moving toward or away from an observer will create horizontal binocular disparities that change coherently over time, the same object will produce different horizontal velocities upon the two retinæ. In principle, these velocity signals (e.g., the output of banks of monocular Reichardt/motion energy detectors) could be compared across the eyes to recover motion through depth. This putative velocity-based source of binocular 3D motion information has been termed the interocular velocity difference (IOVD) cue.

The CD and IOVD cues are naturally conjoined, in that a real moving object will produce both in geometric concordance. In the mathematical sense, the only difference between CD and IOVD is the order of operations in performing horizontal comparisons over time and between the eyes. Note, however, that the order of operations has an important implication for how we think about visual cortex: Computing retinal disparities and tracking them over time is consistent with what was known about primate visual cortex, whereas subtracting two monocular motion signals, one from each eye, is not. The latter is a noncanonical computation because binocularity and motion sensitivity are both thought to emerge exclusively at the same level in the primate visual system (V1), and so there would not be any strictly monocular motion signals to compare for computing an IOVD. The sidebar, *Trials and Tribulations of Cue-Isolating 3D Motion Stimuli*, describes subsequent attempts to isolate the IOVD cue from the CD cue via stimulus design.

Experimentally, the value of isolating these cues is to test whether the visual system fundamentally works with CDs or IOVDs; later in the review, we discuss the evidence demonstrating that both are used. Such cue isolation does not make for ecologically realistic displays, however, as the isolation of either cue requires fairly devilish trickery on the part of the experimenter. Furthermore, the questionable ecological validity of such stimuli is underscored by their inherently retinal conceptions: As opposed to designing stimuli in terms of real object motions and then rendering the appropriate projections, they are typically built at the level of retinal disparities or velocities. This results in several simplifications that we discuss in detail after we introduce a geometric framework for appreciating and synthesizing both realistic and controlled stimuli.

1.2. A Framework for Relating Motion and Depth

The preceding selective discussion does not mean to imply that motion and depth have always been treated in complete isolation. In fact, Gibson's notion of the optical flow field is a fundamental piece of the vision science canon, and it explicitly and inextricably connects motion and depth (Gibson 1979). The optical flow field also allows for the integrated consideration of both object and observer motion. The idea is that any sort of movement in a 3D scene will generate motion in the visible portion of the ambient optic array and thus upon the retina. Object motions create local patterns of velocities, observer motion produces large-field changes in the pattern of velocity, and real-world environments typically produce complex mixtures of these velocities that vary over space and time. Perhaps the only shortcoming of the optical flow field is that it does not consider binocular mechanisms, but it

is perhaps all the more remarkable in having specified how much information is available in a dynamic monocular scene.

Given the qualitative richness of the monocular optic flow field and the fact that we have two eyes, we proffer the binoptic flow field (BFF) as the starting point for thinking about dynamic 3D information and experimental stimuli. It is a nearly complete description of the visual information available from the perspective of the dorsal stream (i.e., ignoring visual features like color/wavelength) and thus can be thought of as a dorsal-stream distillation of the more general plenoptic function (Adelson & Bergen 1991). Indeed, binocular versions of optic flow are perceived especially well (van den Berg & Brenner 1994). By simply incorporating both eyes, the BFF integrates CD, IOVD, and optic flow by completing the binocular geometric relations. This means that it describes the patterns of CD and IOVD as a function of position in the visual field and as a function of object structure and distance in the environment. More practically (but of equal import), it allows one to (*a*) calculate the retinal velocities and stimulus projections corresponding to viable real-world stimuli and, going in the other direction, (*b*) calculate real-world trajectories, cue conflicts, etc., for any previously or commonly used experimental stimuli.

We illustrate a simple example BFF in Figure 1 (see figure caption for additional details). Briefly, panel *a* shows a photo of the retinal images of a realistic model eye viewing a “frozen” flow field from both the left-eye and right-eye positions. If nothing else, this realistic depiction of the proximal stimulus for vision should serve as a reminder that real, biological visual systems accomplish a remarkable feat, given the rather messy input dictated by real optics, vasculature, etc. In panels *b–d*, we show a computer simulation of the same situation. Figure 1*b* shows an overview of the situation. The arrows projecting from the plane show the motion vectors. Figure 1*c* shows the view from behind the eyes (assuming thin lens, 16-mm focal length, 25-mm eye diameter, and the center of rotation is the center of the eye). The retinal projections of the spatial points are shown for five instants in time. In this view, one can appreciate the general increase in retinal velocities with decreasing egocentric distance for an object moving at constant velocity. One can also appreciate that, for nonzero vergence postures, the focus of expansion (FOE) will be at noncorresponding retinal loci (each FOE being, in this case, directly in front of each eye). In Figure 1*d*, we show our conventional representation of the BFF. Here, we first rotate and shift each eye’s flow field so that the ideally corresponding retinal locations are aligned. We then plot the motion vectors for the right and left eyes in this retinal coordinate system. In this representation, the horizontal disparities, vertical disparities, and the corresponding velocity differences across the eyes become evident, as do the associated spatial gradients of these terms.

Below, we first selectively review the perception and neurophysiology of binocular 3D motion processing, highlighting the findings that are of relevance to the expanded geometric framework explicated by the BFF. Then, by considering these findings in the context of the BFF, we motivate choices of future experiments and stimuli. We believe that this framework brings to light the importance of how velocities depend on distance to the observer and position in the visual field and in turn how these dependencies typically produce richer monocular and dichoptic patterns of velocities than have been considered. Understanding

and manipulating both the retinal projections and the actual distance and position of moving objects that create these retinal projections will be of the utmost importance and offer an opportunity for moving visual neuroscience from the study of how the visual system processes simplified retinal inputs to how the visual system infers the dynamic 3D world.

2. BINOCULAR CUES FOR THE PERCEPTION OF 3D MOTION

Deconstructing 3D motion into specific cues can provide important insights into how useful these sources of information are under real-world conditions. The oft-cited real-world example of a cricket player's ability to hit a rapidly approaching ball is illustrative (Cynader & Regan 1978). In this somewhat-constrained scenario, an observer can roughly orient their gaze to the proper location in space and time to maximize the chance of detecting small changes in the object's disparity (Land & McLeod 2000). Still, the player must detect an approaching object in a noisy environment with only approximate knowledge of the spatial and temporal location of the signal. Even small fixation errors could put the target outside of the upper disparity limit for stereopsis (Blakemore 1970). Thus, even though we can and do dissect cues to 3D motion in the laboratory, we should keep in mind that many or all such cues might have to operate in concert for even relatively simple tasks; the notion that the visual system operates on these cues independently should not be taken as a given.

2.1. Evidence for the Use of Changing Disparities

By exploiting slight differences in perspective created by the horizontal offset of the two eyes, the visual system is able to reconstruct a three-dimensional representation of the world around us. The understanding of such binocular computations has been focused on positional stereopsis, in which an object's position in depth is determined from horizontal retinal disparity (Wheatstone 1838). Expanding on the notion that 2D motion could be computed by tracking an object's retinal position over time (e.g., Braddick 1974), binocular 3D motion computations could be performed by tracking an object's disparity-based position in depth over time—the CD cue.

As described earlier, one of the most influential contributions to the study of 3D motion processing was the development of random-dot stereograms by Bela Julesz. The observation that 3D motion percepts can be produced from purely cyclopean signals (Norcia & Tyler 1984) became a driving force behind the belief that the CD cue is the primary cue to 3D motion (Cumming & Parker 1994, Regan & Gray 2009). Cumming (1995) showed that stereoacuties for static and dynamic disparities exhibit similar detection thresholds and share a common dependence on visual eccentricity and mean disparity pedestal. Therefore, 3D motion detected by a changing disparity-based mechanism is likely to be built from the same binocular disparity computations that support static stereopsis.

Results from motion adaptation paradigms also support the notion that disparity-based computations are indeed used in 3D motion processing. CD-isolating stimuli lack any coherent monocular motion information, yet motion aftereffects have been found following adaptation to CD-isolating stimuli (Czuba et al. 2011, 2012), and cue-specific 3D motion aftereffects have even been observed following simultaneous adaptation to cue-isolating stimuli moving in opposite 3D directions (Joo et al. 2016). Although definitive

electrophysiological support for the CD mechanism remains elusive, a physiologically plausible model for a CD mechanism has been proposed by Peng & Shi (2010, 2014). By taking existing models for static disparity energy computations (Ohzawa et al. 1990) as the inputs for canonical 2D motion energy models (Adelson & Bergen 1985, Watson & Ahumada 1985), their CD energy model for 3D motion processing shares many dependencies that have been observed in prior psychophysical studies (Brooks & Stone 2004, 2006).

2.2. Evidence for the Use of Interocular Velocity Differences

An object moving through depth produces not only changing disparities but different monocular velocities in the two eyes as well. In the extreme case—an object on the sagittal plane moving directly toward or away from the observer—equal and opposite velocities of horizontal motion are projected on the two eyes. By encoding these interocular velocity differences (IOVDs) in corresponding regions of the two retinæ, the visual system could in principle compute 3D directions of motion using a purely velocity-based mechanism.

Although interest in the velocity-based cue has increased in recent years, the notion of IOVDs was present in some of the earliest studies of binocular 3D motion processing (Zeki 1974a). Because of the theoretical difficulty of truly isolating the IOVD cue (i.e., having binocularly matched local velocities without any possible disparity signal), many studies were limited to inferences based on sensitivity differences between CD-isolating stimuli and stimuli that contained both IOVD and CD cues (e.g., Portfors-Yeomans & Regan 1996). These studies tended to favor a disparity-based mechanism (but see Section 2.4). However, as techniques for creating strongly IOVD-biased stimuli improved, so did the apparent contributions of velocity-based mechanisms (Czuba et al. 2010, Fernandez & Farell 2005, Rokers et al. 2008, Sheliga et al. 2016). These methods have facilitated a greater understanding of the relative utility of IOVD versus CD cues and highlighted an increasing need for better models of binocular motion processing (Baker & Bair 2016).

2.3. Evidence for Separate Velocity- and Disparity-Based Circuits

The presence of (a) disparity selectivity as early as V1 (Poggio et al. 1985) and (b) interocular transfer of monocular motion aftereffects (Blakemore & Campbell 1969) indicates that the monocular visual pathways are combined at relatively early stages of visual processing. It follows that all cortical motion signals should be inherently cyclopean (i.e., lacking any eye-of-origin information). Thus, current models of binocular processing posit that when signals reach MT, most if not all neurons are driven by signals from both eyes (Carney & Shadlen 1993, Maunsell & Van Essen 1983).

In contrast to this cyclopean motion stream assumption, the presence of a late-stage IOVD computation is supported by evidence that the visual system is capable of performing IOVD computations on monocular pattern motion signals. Using a set of binocularly paired or unpaired monocular stimuli comprising many small drifting Gabor patches (termed psuedoplaids), 3D motion percepts have been shown to be consistent with comparisons of pattern motion signals arising from monocular intersection-of-constraints operations (Adelson & Movshon 1982). Crucially, however, the monocular component motion

integration can be done over areas much larger than V1 receptive fields (Rokers et al. 2011), and component adaptation effects occur even when the local monocular motion signals are in noncorresponding retinal locations between adaptation and test (Greer et al. 2016).

Additional evidence for a late-stage IOVD computation comes from an adaptation paradigm examining spatial frequency integration in binocular 3D motion processing. By exploiting incomplete interocular transfer of monocular motion aftereffect (MAE) (Mitchell et al. 1975), an illusory IOVD cue can be induced by means of unequal motion aftereffect magnitudes in the two eyes (Brooks 2002b, Fernandez & Farell 2006). Unlike 2D MAEs, 3D components of IOVD-based MAEs were robust to differences in the spatial frequency of adaptation and test stimuli, suggesting independent spatiotemporal integration for 2D and 3D motion processing (Shioiri & Matsumiya 2009).

2.4. Two-Circuit Model of 3D Motion

Although studies over the last decade have found a more significant role for the IOVD cue, differences in stimuli and tasks may play a role in the estimated importance of the two cues. For instance, elements of 3D motion that an experimenter considers important to performance (e.g., velocity dependence), as well as the chosen metric of sensitivity, can influence the estimated relative contributions. The CD cue has typically been found to dominate for relatively small, central stimuli (typically 7° eccentricity) and in tasks measuring minimum detection or displacement thresholds (Cumming & Parker 1994, Harris & Rushton 2003, Portfors-Yeomans & Regan 1996). Likewise, a stronger role for IOVDs has been found in studies using larger or more eccentric stimuli and when measurements of direction and speed discrimination (versus detection) are used (Brooks 2002a, Brooks & Stone 2006, Czuba et al. 2010, Rokers et al. 2009). These differences may be more than artefactual in that the visual system might use different sources of binocular information depending on the relative fidelity of the cues in a given situation or the demands of a particular task (Allen et al. 2015).

Individual differences in the ability to perceive 3D motion have enriched the complexity of understanding 3D motion processing. Individual differences have been shown to exist in observers' ability to perceive 3D motion (Regan & Beverley 1979), for specific binocular cues (Nefs & Harris 2010), and across the visual field (Regan et al. 1986). Interestingly, spatial irregularities in 3D motion sensitivity—stereomotion scotomas—are not well predicted by deficits in either lateral motion or static disparity discrimination thresholds (Hong & Regan 1989) and may reflect specific deficits in binocular 3D motion computations (Barendregt et al. 2014, 2016).

2.5. Broader Incorporation of the Binocular Cues

To date, studies of binocular 3D motion have focused on CD and IOVD cues, which is perfectly reasonable given the historical context described earlier. However, the use of simple cue-isolating stimuli can limit the ecological validity of experimental results and, more subtly, can also alter the ways in which information is weighted by the visual system during the actual experiment [see Discussion and Backus et al. (1999)]. Vertical disparities, for example, are an integral geometric feature of binocular vision but have been largely

neglected in studies of binocular motion perception. Yet vertical disparities have been shown to contribute to binocular stereoacuity (Phillipson & Read 2010), to scale disparity information (Rogers & Bradshaw 1993), and to drive reflexive vergence (Mulligan et al. 2013). Perhaps even more surprisingly, vertical IOVDs have been shown to drive not only horizontal but also vertical vergence (Sheliga et al. 2016). Given the geometry of binocular vision, it is relatively straightforward to determine the projections produced on the retinae. The same projective geometry that defines monocular optic flow information (Gibson 1950) also gives rise to the differing monocular signals for velocity- and disparity-based binocular computations—a simple by-product of the two eyes being horizontally offset but otherwise following the same fundamental geometry. The importance of integrating binocular and monocular cues to 3D motion has been proposed as a crucial component to solving the inverse 3D-motion-correspondence problem (Lages & Heron 2010) and is an important component of recognizing the full geometric dependencies inherent to binocular visual perception.

While the understanding of how binocular signals contribute to 3D motion processing provides insight into how motion signals are combined across the two eyes, new questions are raised about the internal frame of reference for 3D motion processing. The information provided by the IOVD and CD cues is inherently egocentric; to be translated into any environmental frame of reference, it must be integrated with other sources of information. Although electrophysiology, neuroimaging, and psychophysical literature suggest that 2D motion is represented in retinotopic coordinates at least through MT (Gardner et al. 2008, Knapen et al. 2009, Van Essen et al. 1981), it is unclear just what a retinotopic representation would mean for a signal that is inherently higher dimensional than retinal coordinates allow (Barendregt et al. 2015). Three-dimensional motion may thus require a representation that is ultimately more closely linked to 3D space and/or motor planning, as opposed to a 2D retinotopic space.

3. BRAIN MECHANISMS FOR 3D MOTION INFORMATION

The neural processing of frontoparallel motion in the primate brain has been extensively studied, and much is known about the corresponding pathways [see Born & Bradley (2005) and Maunsell & Newsome (1987) for excellent reviews]. In primates, motion processing arises in V1 and many of these direction-selective neurons then project to area MT (Movshon & Newsome 1996), where more than 90% of cells are direction selective (Zeki 1974a). MT receives the majority of its input directly from V1 layer 4 but also gets input from cortical areas V2 and V3, as well as from subcortical routes (Born & Bradley 2005). Not surprisingly, area MT has thus been attributed as having a key role in the processing of motion.

3.1. Initial Characterization of Motion, Depth, and Optic Flow Encoding

In general, neurons in MT are selective for motion in the frontoparallel plane, with a typical tuning bandwidth of $\sim 90^\circ$ around a neuron's preferred motion direction (Britten & van Wezel 1998). These neurons respond to a broad range of speeds with a speed tuning curve that can be approximated by a log-Gaussian function (Nover et al. 2005). MT itself has a

retinotopic organization with receptive field sizes increasing with eccentricity of the receptive field center (with the receptive field diameter approximately equal to eccentricity). MT neurons have, however, been shown to be selective for depth as well (DeAngelis & Uka 2003). Maunsell & Van Essen (1983) reported two-thirds of MT neurons were selective for binocular disparity, both horizontal and vertical. DeAngelis et al. (1998) found that disparity tuning was organized into clusters of neurons with similar preferred disparity and demonstrated via microstimulation that these neurons were used in disparity judgments. It has recently been shown that the disparity tuning of MT neurons arises from input from the V2–V3 pathway to MT (Ponce et al. 2008), rather than from the direct input from V1. MT neurons have also been shown to be selective for motion parallax during self-motion (Nadler et al. 2008).

Although these features appear helpful for encoding frontoparallel motion at different depth planes, they do not directly encode a motion trajectory in 3D space (e.g., toward the observer). For this, a velocity along the depth axis would need to be encoded using either monocular (optic flow) or binocular (CD/IOVD) 3D motion cues. MT neurons have been regarded as unlikely to be a final stage for the representation of optic flow, which typically involves very large stimuli with complex directional patterns, relative to the moderate receptive field size and linear directional tuning of MT (Britten 2008, Lappe 2000). Two areas receiving strong input from area MT, MST and the ventral intraparietal area (VIP), however, appear to explicitly encode such optic flow motion (e.g., Bremmer et al. 2002, Duffy & Wurtz 1991, Lagae et al. 1994, Schaafsma & Duysens 1996). Although the visual motion properties of both areas appear very similar at first, there are larger differences regarding interactions with other modalities and their potential roles in motion processing.

The majority of cells in area MST are disparity tuned (90%), of which 40% show a pattern of selectivity consistent with motion parallax (Roy & Wurtz 1990). MST is also selective to vestibular input (Bremmer et al. 1999, Duffy 1998) and plays a causal role in some visual and vestibular heading judgment tasks (Britten & van Wezel 1998, 2002; Gu et al. 2012). MST thus seems to be a critical stage in the processing of self-motion. VIP, like MST, is selective for visual and vestibular input (Bremmer et al. 2002, Schlack et al. 2002) but is also responsive to auditory cues (Schlack et al. 2005) and touch (Duhamel et al. 1998), with spatial receptive fields that often overlap for different modalities and with visual receptive fields that are reported to be partially in a head-centric reference frame (Duhamel et al. 1997) [but see also Chen et al. (2014)]. This set of multisensory representations from auditory to touch has led to a proposed role of encoding near peripersonal space (Duhamel et al. 1998). Unlike MST, inactivation of VIP does not lead to a reduction in sensitivity in a visual or vestibular heading judgment task (Chen et al. 2016; but see Zhang & Britten 2011). Thus, despite both areas having coarsely similar physiological properties, MST appears well suited to processing optic flow derived from self-motion, whereas VIP appears to extract aspects of optic flow that would occur as an object approached or receded from peripersonal space.

3.2. Neural Encoding of 3D Motion

Because of the nature of the binocular cues and the prevailing view of a strictly cyclopean extrastriate cortex, IOVD- or CD-encoding neurons were not sought experimentally as often as were their monocular/cyclopean counterparts sensitive to optic flow. One early study reported that 20% of cells in cat areas 17 and 18 responded best when oscillating bars moved in opposite directions in the two eyes (Cynader & Regan 1978, 1982). A follow-up study (Cynader & Regan 1982) yielded similar results and found that such cells were also insensitive to variations in disparity. Other work in later extrastriate areas of the cat also revealed 3D tuning, but the direct relevance of such neurons to the primate visual system was unclear (Toyama et al. 1985, 1986). Poggio & Talbot (1981) reported similar results for a small subgroup of about 3% of neurons in macaque area V2. Anecdotal reports for 3D tuning in monkey MT (by virtue of opposite direction preference in the two eyes) came from Zeki (1974b). Although the first extensive and systematic study of such CD/IOVD signals in MT by Maunsell & Van Essen (1983) did not report any neurons with opposing direction preference in the two eyes, they did find some neurons with responses that could be interpreted as encoding motion off the frontoparallel motion axis. However, this off-axis tuning disappeared when the stimulus was placed in the neurons' preferred depth plane.

Thus, despite the early reports, MT was not thought to encode motion in depth directly but rather to code features (frontoparallel motion, disparity) perhaps used to build a representation of motion through depth extracted by later areas. However, the mixed physiological and psychophysical evidence for the use of IOVD (and CD) signals did prompt some human neuroimaging studies, which found CD-selective (Likova & Tyler 2007) and then CD- and IOVD-driven (Rokers et al. 2009) responses in and around human MT.

The accumulation of psychophysical evidence for velocity-based computations in binocular 3D motion processing, combined with nascent neuroimaging evidence for the involvement of human MT+, motivated recent reexaminations of how 3D motion information might be encoded by single neurons in monkey MT. Two independent and complementary studies recently examined this issue. Czuba et al. (2014) found that the majority of neurons in MT (70%) encode information about 3D direction of motion on the basis of binocular cues. Approximately one-third of those (17% of the total) preferred motion trajectories directly toward or away from the observer (note that this is actually a neural overrepresentation because, at most reasonable viewing distances, the distance between the eyes subtends far less than 17% of the horizontal visual field). The measured binocular tuning exhibited quantitative deviations from additive (linear) predictions derived from responses to monocular motion components, suggesting a contribution specific to binocular motion computations. Because of this nonlinear interaction, binocular stimulation with the same motion direction in both eyes yielded only minimal responses in some of these neurons and could thus explain why such cells were only sparsely identified in earlier studies (being possibly ignored as unresponsive).

A complementary paper by Sanada & DeAngelis (2014) measured selectivity in primate MT for 3D motion using IOVD-biased and CD-isolating stimuli. With both cues present, they found selectivity for 3D motion in ~58% of the recorded population. Whereas 56%

responded to the IOVD-biased stimulus, only 10% responded to the CD-isolating one. Moreover, response to the CD stimulus was comparatively weak, and all but one of the CD-selective neurons were also selective for either the IOVD stimulus or the combined IOVD and CD cue stimulus.

Although these results clarify the existence of 3D motion information in the responses of MT neurons, they raise the question of how MT comes to encode 3D motion. A recent study revealed that many MT neurons have different preferred directions across subregions within the receptive field (Richert et al. 2013). These appeared to be approximately in opposite directions and might contribute either to encoding an FOE in optic flow, IOVDs, or both. Likewise, although Sanada & DeAngelis (2014) demonstrated both IOVD and CD signals in MT, their careful quantification did not directly answer whether the two binocular cues represent a merged 3D direction signal or rather separate IOVD and CD subcircuits within MT. In a series of adaptation studies using both fMRI and perceptual motion aftereffects, Joo et al. (2016) found little to no adaptation transfer between cues—suggesting mostly separate subcircuits for IOVDs and CDs within human MT, but direct electrophysiological studies specifically addressing this question are needed. Finally, recent modeling work has suggested a nonintuitive relation between pattern motion selectivity and 3D direction sensitivity (Baker & Bair 2016), so empirical tests in the context of evolving computational models may provide important insights into the fundamental computations performed by MT.

On the whole, then, it seems that MT carries information about many of the cues that might be used to compute 3D motion. In fact, at least for small objects, the population response in MT appears to carry all information necessary to decode relative motion trajectories through depth, which includes the current object location in 3D space: eccentricity (receptive field) and depth (disparity, motion parallax) as well as the 3D motion vector (CD, IOVD, and potentially FOE detectors). This highlights MT's potential role as more than just a frontoparallel motion hub but rather as a general encoder of motion through space. Of course, the fundamental role of MT in 3D motion perception is by no means entirely clear. It may still represent building blocks, but these building blocks correspond clearly to 3D motion (CD and IOVD signals), in addition to carrying retinal motion signals and static binocular disparities. These are perhaps used by later areas to compute something like a BFF or at least to fuse the velocity-based and disparity-based motion signals into a single representation of 3D direction. Furthermore, although causal perturbations of MT activity are known to have large and specific effects on frontoparallel motion perception and represent some of the most crucial evidence tying MT to motion perception (Newsome & Paré 1988), similar studies using 3D motion are needed.

4. DISCUSSION

The study of 3D motion, by its very nature, can potentially provide a unifying framework, extending classical approaches for decomposing visual information and mechanisms into retinotopic or image-based constituents. The hope is that this extended framework incorporating environmental variables explicitly will help further the understanding of how the nervous system makes inferences about the external world using information projected

onto the sensory periphery. In this section, we attempt to build a conceptual segue from the retino-centric approach of most prior work to this more ecological framework that can perhaps help guide future work.

The majority of the work we reviewed focused on dissecting classical visual features within an explicit or implicit retino-centric framework. The specific focus on distinguishing IOVD and CD cues is most glaring. This body of work has revealed distinct mechanisms for velocity-based and disparity-based 3D motion information, and moreover, these mechanisms have different functional domains. The encoding of similar information into distinct sensory channels is by no means unique of course. The most obvious example is the presence of rods and cones; although they both transduce light, they actually function as largely independent systems with distinct properties that can be revealed psychophysically. In the case of 3D motion, the purview of the CD system seems to be low speeds in the central visual field, whereas that of the IOVD system seems to be higher speeds over much of the visual field (Czuba et al. 2010).

Although certainly foundational, the stimuli used in much of this prior work spanned but a small slice of binocular stimulus space and often contained unintentional cue conflicts (e.g., large fields with no vertical disparities). We propose that although the cue-isolating perspective has indeed taught us much about the structure of the visual motion processing pathways, it is perhaps time to begin thinking about amore ecologically informed approach to guide future work. Without this complementary perspective, the continued focus on IOVD-CD distinction may spiral into triviality for two related reasons. First, the stimuli used to isolate CDs and IOVDs are intentionally constrained to render 3D motion with only one of two binocular sources of information, but as a result, they differ along a number of nuisance dimensions. Thus, inferences about IOVD versus CD distinctions must always be interpreted in light of other differences between the stimuli. And although some protocols (e.g., cross-adaptation) and convergent evidence across laboratories provide a fairly compelling picture of mostly distinct initial processing, we should bear in mind that under natural conditions, IOVDs and CDs almost always co-occur. Second, and perhaps more importantly, although these signals are distinguishable, they both appear to flow through MT. It seems unlikely that later stages, which could presumably read out MT for estimates of 3D motion, would continue to keep velocity-based and disparity-based representations entirely separate. It would be far more reasonable for a readout mechanism simply to treat such signals, finally, as estimates of 3D velocity. In other words, it seems likely that later stages of the visual system would attempt to encode the actual environmental speeds of objects and the observer's movement through the 3D world (e.g., distance scaling of size, disparity, and motion parallax)—computations we know dispiritingly little about.

Environmental inferences are vital, as having a nervous system that merely decomposes the retinal stimulation is unlikely to tell us much about the actual world that is useful for metrically accurate actions in response to the physical environment. Of course, the abstraction of motion or temporal change is a valuable starting point for a visual system—it gives an animal the “burglar alarm” visual system discussed earlier, for example—but, ultimately, as the evolutionary arms race migrated to dry land and clear air, the winning animals were primarily those that could use vision to navigate the 3D environment and

interact with other animals in it. This, in turn, required a translation of sensory input into more elaborate behaviors, the planning of which would seem to require a common and fairly accurate representation of a dynamic 3D world. The BFF offers a tractable geometric framework for designing stimuli and experiments that begins with a 3D model of the physical world and that then considers, and perhaps dissects, the dynamic retinal projections for both eyes. There are two pragmatic advances that immediately fall out of this perspective. First, it allows for generation of stimuli with the position and velocity specified in 3D-world coordinates and units. For example, considering the structure within the BFF, one can appreciate that the constant retinal speeds used in most 3D motion stimuli correspond to steeply accelerating real-world motions, and the steepness of the acceleration depends on the viewing distance. Second, the consideration of realistic retinal projection forces one to realize that setting various cues to zero does not mean that they are not present in the stimulus but instead that the stimulus either contains cue conflict or that it has a different environmental interpretation than the experimenter perhaps intended. For example, 3D motion stimuli that do not include size changes are not simply avoiding the study of looming (changing retinal size) contributions but are in fact forcing the vertical velocities and horizontal and vertical gradients to zero, which would only be consistent with a very “just so” nonrigid object or navigation through a geometrically serendipitous nonrigid environment.

Although the BFF is useful for dissecting and constructing experimental stimuli, it is also valuable in providing more conceptual insights about the relationships between environmental and retinal variables. These are revealed by manipulating 3D positions instead of retinal variables and then computing the required retinal variables using the BFF; three examples of this are shown in Figure 2 (see caption for details). Below, we separately explain two of these sorts of insights, first for the relation between distance and interocular velocity differences for a model cell tuned to different speeds in the same direction in the two eyes, and then for the relation between eccentricity and stimulus direction for a model cell tuned to opposite directions of motion in the two eyes.

The complex relationship between retinal and environmental coordinates can be appreciated in Figure 3, which plots the response of a simulated IOVD neuron (*color*) as a function of 3D direction in environmental coordinates (x axis) and viewing distance (y axis). Figure 3*a* corresponds to an environmental velocity of 50 cm/s (e.g., a toddler’s walking speed). The white lines show contours of equal velocity ratio between the eyes. Note that, at near viewing distances, a quarter of all 3D motion trajectories (and therefore velocity ratios) would hit the observer in the head [see Regan (1993)], whereas, at far distances, very few do (the dashed line transects 10:1 binocular velocity ratios, which corresponds to the simulated neuron’s actual tuning of 10°/s in the left eye and 1°/s in the right). At near viewing distances, the neuron fires most briskly to a stimulus heading just to the right of the right eye, but, at far viewing distances, an object moving at the same environmental speed can excite the neuron only if it is moving basically rightward. Panels *b* and *c* show the neuron’s response to speeds of 200 cm/s (e.g., a brisk, adult walking speed) and 1,200 cm/s (e.g., a thrown ball). Equivelocity contours are shown for each eye for 1°, 5°, 10°, and 25°/s; note the dramatic change with viewing distance (and of course the correspondence to the neural response). The open circles show the intersection of the two eyes’ preferred velocity

contours and hence the peak response of the unit. Figure 3*d* shows the responses of the neuron for the three environmental speeds as a function of viewing distance (distance along the dashed transects in Figure 3*a–c*). To disambiguate changes in neuronal responses resulting from possible changes in motion direction, speed, or viewing distance, the visual system must rely on additional sources of information. By examining at what stage responses of the dorsal stream hierarchy develop invariances to geometric characteristics of the BFF, we can advance models of how the visual system encodes information about not only the patterns of light projected on the retinae but an actionable representation of the dynamics and structure of the 3D world.

For a second example, we consider a toy neuron that might be found in the subpopulation of cells in area MT (roughly 17%) that comprises cells tuned to opposite directions of motion in the two eyes [Czuba et al. 2014 (e.g., figure 2*e–h*)]. Such a neuron encodes what might best be called interocular velocity opponency (IOVO). When one is looking at a directly approaching object in the sagittal plane, the motion will be equal and opposite in both eyes. As detailed above (Figure 3), the response of a simulated cell tuned to different monocular velocities of the same sign to different directions of motion through depth depends strongly on the object distance. In contrast, a simulated cell tuned for opposing velocities of the same magnitude will maintain the peak of its tuning curve for 3D motions corresponding to an object coming directly toward the observer across a range of distances (Figure 4). Further, this response is only minimally affected by the horizontal eccentricity of the object (and cell's receptive field) relative to the observer. Although many other neurons appear to also respond differentially for various forms of IOVDs, the contributions of these different subpopulations to perception are only beginning to be understood (Bonnen et al. 2017)—most typical examples of IOVDs in the literature are actually the special case of IOVO.

To illustrate this geometric invariance, we consider the response of such a unit to objects arranged on a circle around an observer for motions in all directions as shown in Figure 4*a*. This panel depicts five objects, each of which can move in any direction. Superimposed on each object's starting position is a polar plot of the response of a simulated cell at that location (radius) as a function of the direction of the object's motion (angle). In panels *b–d*, the stimulus and response are plotted as a function of environmental direction, whereas in panels *e–g*, the stimulus and response are plotted as a function of egocentric direction. Panel *b* shows the retinal velocities created by each stimulus (*color*) to each eye (left = *solid*, right = *dashed*). Panel *e* shows the same velocities but now plotted relative to the observer; the inset more clearly shows that the projected velocities are of opposite sign in the two eyes when the object moves toward the head. Panels *c* and *f* show the corresponding IOVDs computed from panels *b* and *e*; the staggering of the curves in panel *e* is due to the relative distances of the stimuli to the two eyes. Panels *d* and *g* show the responses of the simulated cells. We can see that the maximal response is always obtained for directions very nearly directed at the nasion (and the small offsets in the curves for the eccentric objects are, again, due to the relative distances to the eyeballs). This illustration should make it clear that so-called IOVO detectors, cells tuned to opposite retinal directions in the two eyes, are very potent for signaling approaching objects that might collide with the observer. In addition to being useful for any estimation of heading, these cells may be especially useful for detecting intercepting objects. As such, it would be interesting to know where these cells project and

how their signals are used. Area VIP, which may play a role in the encoding of the near peripersonal space, appears to be a prime candidate for a region to make use of this information.

In summary, the geometric framework of the BFF motivates careful consideration of what the real-world distal stimuli are and how these stimuli project to the two eyes to create a rich interocular pattern of retinal information. Moreover, the computing power we all now enjoy in our laboratories makes it possible to use the real, 3D, dynamic world as the starting point for stimulus generation while still preserving the level of control over the actual retinal stimulation that we demand. In addition to affecting how studies of encoding are performed (i.e., which stimuli are used, which are considered most basic, and how various cues are isolated), consideration of the BFF should also steer decoding analyses, which typically focus on the value of variables as they fall upon the retina. It seems safe to assume, however, such decoding (read-out) processes must ultimately support inferences about the 3D world, as that is the world in which animals have to survive.

Acknowledgments

We wish to thank Kathryn L. Bonnen for helpful discussions, and Kenneth Latimer for early analyses. L.K.C. and A.C.H. were supported by NIH-NEI-EY020592; A.C.H. was previously supported by NSF CAREER BCS-0748413.

LITERATURE CITED

- Adelson EH, Bergen JR. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*. 1985; 2(2):284–99. [PubMed: 3973762]
- Adelson, EH., Bergen, JR. The plenoptic function and elements of early vision. In: Landy, MS., Movshon, JA., editors. *Computational Models of Visual Processing*. Cambridge, MA: MIT Press; 1991. p. 3–20.
- Adelson EH, Movshon JA. Phenomenal coherence of moving visual patterns. *Nature*. 1982; 300:523–25. [PubMed: 7144903]
- Albright TD. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiol*. 1984; 52(6):1106–30. [PubMed: 6520628]
- Allen B, Haun AM, Hanley T, Green CS, Rokers B. Optimal combination of the binocular cues to 3D motion. *Invest. Ophthalmol. Vis. Sci*. 2015; 56(12):7589–96. <http://doi.org/10.1167/iovs.15-17696>. [PubMed: 26618651]
- Anstis SM. Phi movement as a subtraction process. *Vis. Res*. 1970; 10:1411–30. [PubMed: 5516541]
- Backus BT, Banks MS, van Ee R, Crowell JA. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vis. Res*. 1999; 39(6):1143–70. [PubMed: 10343832]
- Baker PM, Bair W. A model of binocular motion integration in MT neurons. *J. Neurosci*. 2016; 36(24):6563–82. <http://doi.org/10.1523/JNEUROSCI.3213-15.2016>. [PubMed: 27307243]
- Barendregt M, Dumoulin SO, Rokers B. Stereomotion scotomas occur after binocular combination. *Vis. Res*. 2014; 105:92–99. <http://doi.org/10.1016/j.visres.2014.09.008>. [PubMed: 25286416]
- Barendregt M, Dumoulin SO, Rokers B. Impaired velocity processing reveals an agnosia for motion in depth. *Psychol. Sci*. 2016; 27(11):1474–85. <http://doi.org/10.1177/0956797616663488>. [PubMed: 27677897]
- Barendregt M, Harvey BM, Rokers B, Dumoulin SO. Transformation from a retinal to a cyclopean representation in human visual cortex. *Biol. Biol*. 2015; 25(15):1982–87. <http://doi.org/10.1016/j.cub.2015.06.003>. [PubMed: 26144967]
- Blakemore C. The range and scope of binocular depth discrimination in man. *J. Physiol*. 1970; 211(3):599–622. [PubMed: 5501054]

- Blakemore C, Campbell FW. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol.* 1969; 203(1):237–60. [PubMed: 5821879]
- Bonnen, K., Czuba, T., Kohn, A., Cormack, L., Huk, A. Encoding and decoding in populations with non-Gaussian tuning: the example of 3D motion; Presented at Cosyne 2017; Feb. 23–28; Salt Lake City, UT. 2017.
- Born RT, Bradley DC. Structure and function of visual area MT. *Annu. Rev. Neurosci.* 2005; 28:157–89. <http://doi.org/10.1146/annurev.neuro.26.041002.131052>. [PubMed: 16022593]
- Braddick O. A short range process in apparent motion. *Vis. Res.* 1974; 14:519–27. [PubMed: 4423193]
- Bremmer F, Duhamel J-R, Ben Hamed S, Graf W. Heading encoding in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.* 2002; 16(8):1554–68. [PubMed: 12405970]
- Bremmer F, Kubischik M, Pekel M, Lappe M, Hoffmann KP. Linear vestibular self-motion signals in monkey medial superior temporal area. *Ann. N. Y. Acad. Sci.* 1999; 871:272–81. [PubMed: 10372078]
- Britten KH. Mechanisms of self-motion perception. *Annu. Rev. Neurosci.* 2008; 31:389–410. <http://doi.org/10.1146/annurev.neuro.29.051605.112953>. [PubMed: 18558861]
- Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 1992; 12(12):4745–65. [PubMed: 1464765]
- Britten KH, Shadlen MN, Newsome WT, Movshon JA. Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.* 1993; 10(6):1157–69. <http://doi.org/10.1017/S0952523800010269>. [PubMed: 8257671]
- Britten KH, van Wezel RJ. Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nat. Neurosci.* 1998; 1(1):59–63. <http://doi.org/10.1038/259>. [PubMed: 10195110]
- Britten KH, van Wezel RJA. Area MST and heading perception in macaque monkeys. *Cereb. Cortex.* 2002; 12(7):692–701. [PubMed: 12050081]
- Brooks KR. Interocular velocity difference contributes to stereomotion speed perception. *J. Vis.* 2002a; 2(3):2. <http://doi.org/10.1167/2.3.2>.
- Brooks KR. Monocular motion adaptation affects the perceived trajectory of stereomotion. *J. Exp. Psychol. Hum. Percept. Perform.* 2002b; 28(6):1470–82. [PubMed: 12542138]
- Brooks KR, Stone LS. Stereomotion speed perception: contributions from both changing disparity and interocular velocity difference over a range of relative disparities. *J. Vis.* 2004; 4(12):6. <http://doi.org/10.1167/4.12.6>.
- Brooks KR, Stone LS. Stereomotion suppression and the perception of speed: accuracy and precision as a function of 3D trajectory. *J. Vis.* 2006; 6(11):6. <http://doi.org/10.1167/6.11.6>.
- Carney T, Shadlen MN. Dichoptic activation of the early motion system. *Vis. Res.* 1993; 33(14):1977–95. [PubMed: 8249314]
- Chen A, Gu Y, Liu S, DeAngelis GC, Angelaki DE. Evidence for a causal contribution of macaque vestibular, but not intraparietal, cortex to heading perception. *J. Neurosci.* 2016; 36(13):3789–98. <http://doi.org/10.1523/JNEUROSCI.2485-15.2016>. [PubMed: 27030763]
- Chen X, DeAngelis GC, Angelaki DE. Eye-centered visual receptive fields in the ventral intraparietal area. *J. Neurophysiol.* 2014; 112(2):353–61. <http://doi.org/10.1152/jn.00057.2014>. [PubMed: 24790176]
- Cumming BG. The relationship between stereoacuity and stereomotion thresholds. *Perception.* 1995; 24(1):105–14. <http://doi.org/10.1068/p240105>. [PubMed: 7617414]
- Cumming BG, Parker AJ. Binocular mechanisms for detecting motion-in-depth. *Vis. Res.* 1994; 34(4):483–95. [PubMed: 8303832]
- Cynader M, Regan D. Neurones in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. *J. Physiol.* 1978; 274:549–69. [PubMed: 625008]
- Cynader M, Regan D. Neurons in cat visual cortex tuned to the direction of motion in depth: effect of positional disparity. *Vis. Res.* 1982; 22(8):967–82. [PubMed: 7135859]

- Czuba TB, Huk AC, Cormack LK, Kohn A. Area MT encodes three-dimensional motion. *J. Neurosci.* 2014; 34(47):15522–33. <http://doi.org/10.1523/JNEUROSCI.1081-14.2014>. [PubMed: 25411482]
- Czuba TB, Rokers B, Guillet K, Huk AC, Cormack LK. Three-dimensional motion aftereffects reveal distinct direction-selective mechanisms for binocular processing of motion through depth. *J. Vis.* 2011; 11(10):18. <http://doi.org/10.1167/11.10.18>.
- Czuba TB, Rokers B, Huk AC, Cormack LK. Speed and eccentricity tuning reveal a central role for the velocity-based cue to 3D visual motion. *J. Neurophysiol.* 2010; 104(5):2886–99. <http://doi.org/10.1152/jn.00585.2009>. [PubMed: 20881201]
- Czuba TB, Rokers B, Huk AC, Cormack LK. To CD or not to CD: Is there a 3D motion aftereffect based on changing disparities? *J. Vis.* 2012; 12(4):7. <http://doi.org/10.1167/12.4.7>.
- DeAngelis GC, Cumming BG, Newsome WT. Cortical area MT and the perception of stereoscopic depth. *Nature.* 1998; 394(6694):677–80. <http://doi.org/10.1038/29299>. [PubMed: 9716130]
- DeAngelis GC, Uka T. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *J. Neurophysiol.* 2003; 89(2):1094–111. <http://doi.org/10.1152/jn.00717.2002>. [PubMed: 12574483]
- Duffy CJ. MST neurons respond to optic flow and translational movement. *J. Neurophysiol.* 1998; 80(4):1816–27. [PubMed: 9772241]
- Duffy CJ, Wurtz RH. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* 1991; 65(6):1329–45. [PubMed: 1875243]
- Duhamel JR, Bremmer F, Ben Hamed S, Graf W. Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature.* 1997; 389(6653):845–48. <http://doi.org/10.1038/39865>. [PubMed: 9349815]
- Duhamel JR, Colby CL, Goldberg ME. Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *J. Neurophysiol.* 1998; 79(1):126–36. [PubMed: 9425183]
- Fernandez JM, Farell B. Seeing motion in depth using inter-ocular velocity differences. *Vis. Res.* 2005; 45(21):2786–98. <http://doi.org/10.1016/j.visres.2005.05.021>. [PubMed: 16023695]
- Fernandez JM, Farell B. Motion in depth from interocular velocity differences revealed by differential motion aftereffect. *Vis. Res.* 2006; 46(8–9):1307–17. <http://doi.org/10.1016/j.visres.2005.10.025>. [PubMed: 16356530]
- Gardner JL, Merriam EP, Movshon JA, Heeger DJ. Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *J. Neurosci.* 2008; 28(15):3988–99. <http://doi.org/10.1523/JNEUROSCI.5476-07.2008>. [PubMed: 18400898]
- Gibson, JJ. *The Perception of the Visual World*. Boston, MA: Houghton Mifflin; 1950.
- Gibson, JJ. *The Ecological Approach to Visual Perception*. New York: Psychol. Press; 1979.
- Greer D, Joo SJ, Cormack L, Huk A. Global eye-specific motion signal for three-dimensional motion processing revealed through adaptation. *J. Vis.* 2016; 16(12):182. <http://doi.org/10.1167/16.12.182>.
- Gu Y, DeAngelis GC, Angelaki DE. Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *J. Neurosci.* 2012; 32(7):2299–313. <http://doi.org/10.1523/JNEUROSCI.5154-11.2012>. [PubMed: 22396405]
- Harris JM, Rushton SK. Poor visibility of motion in depth is due to early motion averaging. *Vis. Res.* 2003; 43(4):385–92. [PubMed: 12535995]
- Hong X, Regan D. Visual field defects for unidirectional and oscillatory motion in depth. *Vis. Res.* 1989; 29(7):809–19. [PubMed: 2623824]
- Joo SJ, Czuba TB, Cormack LK, Huk AC. Separate perceptual and neural processing of velocity- and disparity-based 3D motion signals. *J. Neurosci.* 2016; 36(42):10791–802. <http://doi.org/10.1523/JNEUROSCI.1298-16.2016>. [PubMed: 27798134]
- Julesz, B. *Foundations of Cyclopean Perception*. Chicago: Univ. Chicago Press; 1971.
- Knapen T, Rolfs M, Cavanagh P. The reference frame of the motion aftereffect is retinotopic. *J. Vis.* 2009; 9(5):16. <http://doi.org/10.1167/9.5.16>.
- Lagae L, Maes H, Raiguel S, Xiao DK, Orban GA. Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST. *J. Neurophysiol.* 1994; 71(5):1597–626. [PubMed: 8064337]

- Lages M, Heron S. On the inverse problem of binocular 3D motion perception. *PLOS Comput. Biol.* 2010; 6(11):e1000999. <http://doi.org/10.1371/journal.pcbi.1000999>. [PubMed: 21124957]
- Land MF, McLeod P. From eye movements to actions: how batsmen hit the ball. *Nat. Neurosci.* 2000; 3(12):1340–45. <http://doi.org/10.1038/81887>. [PubMed: 11100157]
- Land, MF., Nilsson, DE. *Animal Eyes*. 2. Oxford, UK: Oxford Univ. Press; 2012.
- Lappe M. Computational mechanisms for optic flow analysis in primate cortex. *Int. Rev. Neurobiol.* 2000; 44:235–68. [PubMed: 10605649]
- Likova LT, Tyler CW. Stereomotion processing in the human occipital cortex. *NeuroImage.* 2007; 38(2):293–305. <http://doi.org/10.1016/j.neuroimage.2007.06.039>. [PubMed: 17869540]
- Maunsell JHR, Newsome WT. Visual processing in monkey extrastriate cortex. *Annu. Rev. Neurosci.* 1987; 10:363–401. <http://doi.org/10.1146/annurev.ne.10.030187.002051>. [PubMed: 3105414]
- Maunsell JHR, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* 1983; 49(5):1148–67. [PubMed: 6864243]
- Mitchell DE, Reardon J, Muir DW. Interocular transfer of the motion after-effect in normal and stereoblind observers. *Exp. Brain Res.* 1975; 22(2):163–73. [PubMed: 1126412]
- Morgan MJ, Ward R. Conditions for motion flow in dynamic visual noise. *Vis. Res.* 1980; 20(5):431–35. [PubMed: 7414977]
- Movshon JA, Newsome WT. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. *J. Neurosci.* 1996; 16(23):7733–41. [PubMed: 8922429]
- Mulligan JB, Stevenson SB, Cormack LK. Reflexive and voluntary control of smooth eye movements. *Proc. SPIE, Hum. Vis. Electron. Imaging XVIII.* 2013; 8651:86510Z. <http://doi.org/10.1117/12.2010333>.
- Nadler JW, Angelaki DE, DeAngelis GC. A neural representation of depth from motion parallax in macaque visual cortex. *Nature.* 2008; 452(7187):642–45. <http://doi.org/10.1038/nature06814>. [PubMed: 18344979]
- Nefs HT, Harris JM. What visual information is used for stereoscopic depth displacement discrimination? *Perception.* 2010; 39(6):727–44. [PubMed: 20698469]
- Newsome WT, Paré EB. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* 1988; 8(6):2201–11. [PubMed: 3385495]
- Nilsson DE. Eyes as optical alarm systems in fan worms and ark clams. *Philos. Trans. R. Soc. B.* 2006; 346:195–212.
- Nishihara HK. Practical real-time imaging stereo matcher. *Opt. Eng.* 1984; 23(5):536–45. <http://doi.org/10.1117/12.7973334>.
- Norcia A, Tyler C. Temporal frequency limits for stereoscopic apparent motion processes. *Vis. Res.* 1984; 24(5):395–401. [PubMed: 6740960]
- Nover H, Anderson CH, DeAngelis GC. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J. Neurosci.* 2005; 25(43):10049–60. <http://doi.org/10.1523/JNEUROSCI.1661-05.2005>. [PubMed: 16251454]
- Ohzawa I, DeAngelis GC, Freeman RD. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science.* 1990; 249(4972):1037–41. [PubMed: 2396096]
- Peng Q, Shi BE. The changing disparity energy model. *Vis. Res.* 2010; 50(2):181–92. <http://doi.org/10.1016/j.visres.2009.11.012>. [PubMed: 19925822]
- Peng Q, Shi BE. Neural population models for perception of motion in depth. *Vis. Res.* 2014; 101(C): 11–31. <http://doi.org/10.1016/j.visres.2014.04.014>. [PubMed: 24814174]
- Phillipson GP, Read JCA. Stereo correspondence is optimized for large viewing distances. *Eur. J. Neurosci.* 2010; 32(11):1959–69. <http://doi.org/10.1111/j.1460-9568.2010.07454.x>. [PubMed: 20955471]
- Poggio GF, Motter BC, Squatrito S, Trotter Y. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vis. Res.* 1985; 25(3):397–406. [PubMed: 4024459]

- Poggio GF, Talbot WH. Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J. Physiol.* 1981; 315:469–92. [PubMed: 7310720]
- Ponce CR, Lomber SG, Born RT. Integrating motion and depth via parallel pathways. *Nat. Neurosci.* 2008; 11(2):216–23. <http://doi.org/10.1038/nn2039>. [PubMed: 18193039]
- Portfors-Yeomans C, Regan D. Cyclopean discrimination thresholds for the direction and speed of motion in depth. *Vis. Res.* 1996; 36(20):3265–79. [PubMed: 8944286]
- Regan D. Binocular correlates of the direction of motion in depth. *Vis. Res.* 1993; 33(16):2359–60. [PubMed: 8273301]
- Regan D, Beverley KI. Binocular and monocular stimuli for motion in depth: changing-disparity and changing-size feed the same motion-in-depth stage. *Vis. Res.* 1979; 19(12):1331–42. [PubMed: 532101]
- Regan D, Erkelens CJ, Collewijn H. Visual field defects for vergence eye movements and for stereomotion perception. *Invest. Ophthalmol. Vis. Sci.* 1986; 27(5):806–19. [PubMed: 3700030]
- Regan D, Gray R. Binocular processing of motion: some unresolved questions. *Spatial Vis.* 2009; 22(1):1–43. <http://doi.org/10.1163/156856809786618501>.
- Richert M, Albright TD, Krekelberg B. The complex structure of receptive fields in the middle temporal area. *Front. Syst. Neurosci.* 2013; 7:2. <http://doi.org/10.3389/fnsys.2013.00002>. [PubMed: 23508640]
- Rogers BJ, Bradshaw MF. Vertical disparities, differential perspective and binocular stereopsis. *Nature.* 1993; 361:253–55. <http://doi.org/10.1038/361253a0>. [PubMed: 8423851]
- Rokers B, Cormack LK, Huk AC. Strong percepts of motion through depth without strong percepts of position in depth. *J. Vis.* 2008; 8(4):6. <http://doi.org/10.1167/8.4.6>.
- Rokers B, Cormack LK, Huk AC. Disparity- and velocity-based signals for three-dimensional motion perception in human MT+ *Nat. Neurosci.* 2009; 12(8):1050–55. <http://doi.org/10.1038/nn.2343>. [PubMed: 19578382]
- Rokers B, Czuba TB, Cormack LK, Huk AC. Motion processing with two eyes in three dimensions. *J. Vis.* 2011; 11(2):10. <http://doi.org/10.1167/11.2.10>.
- Roy J-P, Wurtz RH. The role of disparity-sensitive cortical neurons in signalling the direction of self-motion. *Nature.* 1990; 348:160–62. [PubMed: 2234078]
- Sanada TM, DeAngelis GC. Neural representation of motion-in-depth in area MT. *J. Neurosci.* 2014; 34(47):15508–21. <http://doi.org/10.1523/JNEUROSCI.1072-14.2014>. [PubMed: 25411481]
- Schaafsma SJ, Duysens J. Neurons in the ventral intraparietal area of awake macaque monkey closely resemble neurons in the dorsal part of the medial superior temporal area in their responses to optic flow patterns. *J. Neurophysiol.* 1996; 76(6):4056–68. [PubMed: 8985900]
- Schlack A, Hoffmann K-P, Bremmer F. Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.* 2002; 16(10):1877–86. [PubMed: 12453051]
- Schlack A, Sterbing-D'Angelo SJ, Hartung K, Hoffmann K-P, Bremmer F. Multisensory space representations in the macaque ventral intraparietal area. *J. Neurosci.* 2005; 25(18):4616–25. <http://doi.org/10.1523/JNEUROSCI.0455-05.2005>. [PubMed: 15872109]
- Shadlen MN, Newsome WT. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* 2001; 86(4):1916–36. [PubMed: 11600651]
- Sheliga BM, Quaia C, FitzGibbon EJ, Cumming BG. Human short-latency ocular vergence responses produced by interocular velocity differences. *J. Vis.* 2016; 16(10):11. <http://doi.org/10.1167/16.10.11>.
- Shioiri S, Matsumiya K. Motion mechanisms with different spatiotemporal characteristics identified by an MAE technique with superimposed gratings. *J. Vis.* 2009; 9(5):30.
- Toyama K, Komatsu Y, Kasai H, Fujii K, Umetani K. Responsiveness of Clare-Bishop neurons to visual cues associated with motion of a visual stimulus in three-dimensional space. *Vis. Res.* 1985; 25(3):407–14. [PubMed: 4024460]
- Toyama K, Komatsu Y, Kozasa T. The responsiveness of Clare-Bishop neurons to motion cues for motion stereopsis. *Neurosci. Res.* 1986; 4(2):83–109. [PubMed: 3808487]

- van den Berg AV, Brenner E. Why two eyes are better than one for judgements of heading. *Nature*. 1994; 371(6499):700–2. <http://doi.org/10.1038/371700a0>. [PubMed: 7935814]
- Van Essen DC, Maunsell JH, Bixby JL. The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization. *J. Comp. Neurol.* 1981; 199(3):293–326. <http://doi.org/10.1002/cne.901990302>. [PubMed: 7263951]
- Watson AB, Ahumada AJ. Model of human visual-motion sensing. *J. Opt. Soc. Am. A*. 1985; 2(2): 322–41. [PubMed: 3973764]
- Wheatstone C. Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philos. Trans. R. Soc.* 1838; 128:371–94.
- Williams DW, Sekuler R. Coherent global motion percepts from stochastic local motions. *Vis. Res.* 1984; 24:55–62. [PubMed: 6695508]
- Zeki SM. Cells responding to changing image size and disparity in the cortex of the rhesus monkey. *J. Physiol.* 1974a; 242(3):827–41. [PubMed: 4449056]
- Zeki SM. Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. *J. Physiol.* 1974b; 236(3):549–73. [PubMed: 4207129]
- Zhang T, Britten KH. Parietal area VIP causally influences heading perception during pursuit eye movements. *J. Neurosci.* 2011; 31(7):2569–75. [PubMed: 21325524]

TRIALS AND TRIBULATIONS OF CUE-ISOLATING 3D MOTION STIMULI

Much research has focused on crafting stimuli that isolate the changing disparity (CD) and interocular velocity difference (IOVD) cues. Dynamic random-dot stereograms provided initial proof of concept that changing disparities could, themselves, support percepts of motion through depth. By modulating the horizontal retinal disparities over time in a dynamically updated random-dot pattern, such displays contain no coherent monocular motions (and hence no IOVDs); referring to the realistic binocular geometry in Figure 1*d* also reveals that they of course also lack the pattern of vertical disparities that a rigid and flat object would generate. IOVDs are trickier to isolate, as dichoptic motions will typically generate dynamic patterns of disparity. Spatially uncorrelated dot patterns moving in opposite directions provide some leverage but of course still contain (uncontrolled) occurrences of CDs. Sparse, anticorrelated displays reflect a refinement of this approach, squelching the disparity signals by inverting the contrast polarity between the two eyes; however, potential second-order signals remain. More elegant stimuli come closer to IOVD purification by playing off the known (small) sizes of V1 receptive fields (which are classically conceived of as the point of binocular combination) and by generating dichoptic spatiotemporal quadrature patterns that have 180° interocular phase relations (and perfectly ambiguous patterns of disparity change).

Figure 1.

The binoptic flow field (BFF). (a) Photograph (taken from the Scheiner/Descartes perspective) of the projection of a BFF generated by points on a frontoparallel plane on the retina of a realistic model eye in the left (*left, green-tinted projections*) and right (*right, red-tinted projections*) eye positions. Fixation was on the center of the stimulus at its initial position. The visual system faces the important challenge of converting this proximal stimulus into veridical estimates of self-motion and object motion in the external world. (b) An overview of a computer simulation of a situation similar to that shown in panel a. The eyeballs are 3.25 cm to the left and right of the origin, and the visual axes of the right and left eyes are shown by the red and green lines, respectively. The yellow vectors show the motion of points fixed to a plane moving along the z axis. (c) The retinal projections of the points in panel b shown at five uniformly spaced times during the motion. The eyeballs were modeled as 2.5-mm spheres with centers of rotation at the centers of the spheres and with a 16-mm focal-length thin lens. Fixation was on the center of the stimulus at time zero. The eyeballs and projections are rendered from directly behind them—the Scheiner/Descartes view as in panel a. (d) The retinal projections of the motion vectors shown in panel b, after rotating and shifting to align idealized corresponding retinal points.

Figure 2.

Binoptic flow fields (BFFs) generated as in Figure 1*b–d*. The top subpanels show overviews of the three situations: (*a*) an observer walking over a ground plane, (*b*) an observer walking through a 2-m diameter tunnel, and (*c*) an observer walking to a door opening into a 3-m deep room. The midpoint between the eyes is at (0, 0, 0), and fixation is at eye height at the farthest point in each environment. The visual axes of the right eye are shown by the red lines. In panel *c*, the front wall itself—the wall containing the open door—is not shown for clarity but the corresponding motion vectors are; the motion vectors shown on the floor and back wall are those visible through the doorway. The bottom subpanels show the corresponding BFFs; note that, because these are retinal projections, the lower visual fields correspond to positive *y* values. In panel *c* (*bottom*), the top of the doorway is quite obvious at around *y* = −1 mm. We think that these flow fields make it rather clear that describing depth motion in terms of either changing horizontal disparities or horizontal velocity difference is somewhat impoverished.

Figure 3.

Response of a simulated middle temporal area (MT) neuron to real-world object motion: unequal tuning for rightward motion in each eye. A simple instantiation of an interocular velocity difference (IOVD) neuron is composed of a different speed preference in the left and right eye ($10^\circ/\text{s}$ and $1^\circ/\text{s}$, respectively), which corresponds to a 3D motion trajectory slightly to the right of the observer's head. (a–c) Each panel simulates a fixed object speed equivalent to that of, for example, (a) an approaching toddler, (b) a brisk walk, and (c) a ball tossed from 10 m away (50, 200, and 1,200 cm/s, respectively). Response (spikes per second, colormap) is plotted as a function of viewing distance (y axis) and simulated 3D motion trajectory (x axis; within the horizontal plane). Viewing distances range from a few centimeters (3.25 cm; one-half interpupillary distance) to 5 m. Contours of equal velocity ratio (white dashed lines) are independent of stimulus speed, and thus are shown only in panel a. Black dashed lines indicate the particular contour corresponding to the binocular tuning of the simulated neuron (10:1). Equivelocity contours (left eye, green; right eye, red) for select monocular velocities are overlaid on the response surface and highlight the influence of viewing distance on projected monocular velocity. The open circles show the intersection of the two contours corresponding to the neuron's preferred velocities in each eye (and hence necessarily also lie on the preferred ratio contour) and thus correspond to the peak response of the cell. (d) Simulated response to each object speed on the preferred ratio contour (black dashed lines in panels a–c) as a function of viewing distance. Additional information from the binoptic flow field is necessary to disambiguate responses based on IOVD information alone. Abbreviations: A, away; L, left; R, right; T, toward.

Figure 4.

Response of simulated middle temporal area (MT) neurons to real-world object motion: equal but opposite tuning for motion in each eye. (a) Five example objects are located on a circle around an observer located at (0, 0), with the eyes marked by asterisks. For each object, a cell's response is shown in polar coordinates for motion to all directions. (b–d) In the left column, the stimulus and response are plotted as a function of environmental direction (icons around the polar plot correspond to icons on the x axes). (e–g) In the right column, these are plotted as a function of egocentric direction, the space in which a theoretical collision-detector such as our simulated cell would operate. The rows plot (b, e)

retinal velocities, (c , f) IOVDs, and (d , g) the responses of the cells. The direction of the cells' peak response is largely invariant to stimulus eccentricity. Abbreviations: A, away; IOVD, interocular velocity difference; IOVO, interocular velocity opponency; L, left; OV, ocular velocity; R, right; T, toward.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript