The neural basis of depth perception from motion parallax

HyunGoo R. Kim1,†, Dora E. Angelaki2,3 and Gregory C. DeAngelis1

1Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, NY 14627, USA
2Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA
3Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA

†Present address: Department of Molecular and Cell Biology, Harvard University, Cambridge, MA 02138, USA.

In addition to depth cues afforded by binocular vision, the brain processes relative motion signals to perceive depth. When an observer translates relative to their visual environment, the relative motion of objects at different distances (motion parallax) provides a powerful cue to three-dimensional scene structure. Although perception of depth based on motion parallax has been studied extensively in humans, relatively little is known regarding the neural basis of this visual capability. We review recent advances in elucidating the neural mechanisms for representing depth-sign (near versus far) from motion parallax. We examine a potential neural substrate in the middle temporal visual area for depth perception based on motion parallax, and we explore the nature of the signals that provide critical inputs for disambiguating depth-sign.

1. Introduction

Humans and animals frequently move within their environments. Many behaviours that are essential for survival (e.g. foraging, fighting and fleeing), as well as behaviours for social interaction and entertainment (e.g. shaking hands or playing tennis), involve interacting with objects in three-dimensional space while we are moving in the world. Accurate perception of depth during self-motion is critical for success in many such tasks: for example, a lion will decide whether to chase a deer based on the distance between them, and a tennis player will stop running if the ball is not likely to be within reach.

The brain makes use of a variety of cues to estimate depth. These include pictorial depth cues that are present in a single static image of a scene, such as occlusion, relative size, perspective, shading, texture gradients and blur [1,2]. Although such pictorial cues are valuable in interpreting three-dimensional scene structure, they generally do not provide precise quantitative information about depth. Additional powerful depth cues arise when a scene is viewed from multiple vantage points (figure 1). Binocular disparity cues arise because the two eyes are separated horizontally, and provide information about depth [3,4]. Additionally, when an observer translates through the environment, motion parallax cues also provide a powerful source of depth information.

Motion parallax refers to the difference in image motion between objects at different depths [1]. Although some literature considers motion parallax induced by object motion in a scene (e.g. [5]), we focus here on motion parallax that is generated by translation of an observer relative to the scene (i.e. observer-induced motion parallax). It is worth noting that motion parallax and binocular disparity cues both provide quantitative information about depth because they arise from similar geometry (figure 1; see also [1]). For example, when an observer with the right eye closed translates to the right through one inter-ocular distance (figure 1a), the retinal displacement of an object over time matches the binocular disparity that would occur in the absence of self-motion (figure 1b).
motion parallax cues generally co-occur with other depth cues such as binocular disparity and pictorial cues. Thus, it was necessary to isolate motion parallax cues and also to link them with observer movement. Building upon the successes of using random-dot stimuli to study stereopsis [3], Rogers & Graham developed random-dot kinematograms that simulated corrugated surfaces in depth while eliminating other depth cues [18,20]. In addition, they designed a customized display apparatus that synchronized the motion of random dots with the observer’s head movements. With these advances, subjects reliably perceived three-dimensional structure in random-dot kinematograms viewed monocularly [18].

Critically, without extra-retinal inputs related to head or eye movements, the perceived depth-sign (i.e. near versus far, relative to the plane of fixation) of motion parallax stimuli can be ambiguous [18,21,22]. In Rogers & Graham’s [18] pioneering study, observers translated their head side to side while maintaining visual fixation on a world-fixed target by counter-rotating their eyes. Thus, in principle, signals related to either head translation or eye rotation might have provided the necessary extra-retinal inputs for perception of depth-sign from motion parallax. It was later demonstrated that eye movement signals, in fact, provide a sufficient input to disambiguate depth-sign [23–25]. In one manipulation, Nawrot [24] moved the fixation target with the head such that no overt eye movements were required. Interestingly, subjects perceived inverted depth-sign, consistent with the action of a smooth eye movement command signal that serves to cancel the translational vestibulo-ocular reflex (tVOR). This clearly dissociates perception of depth-sign from head translation. It was possible that retinal slip due to imperfect eye movements might be used to infer depth-sign, however, Najj & Freeman [25] demonstrated that this was not the case. These and subsequent studies have demonstrated clearly that smooth eye movement command signals provide an input that can disambiguate depth-sign based on motion parallax in humans (e.g. [8,26,27]).

It is worth noting that some studies have shown that depth perception from motion parallax depends on the relationship between retinal motion velocity and head velocity [28–31] or viewing distance [9]. When subjects fixate a world-fixed target during head translation, eye velocity is systematically related to both head velocity and viewing distance. Thus, these findings are compatible with the idea that smooth eye movements provide a critical input [26,27].

3. Neural correlates of depth perception from motion parallax

Establishing neural correlates of depth perception based on motion parallax has also posed major challenges, and substantial progress has been made in animal models only within the past decade or so. Schiller and colleagues took an important step by training macaque monkeys to discriminate among surface patches that differed in depth defined by motion parallax. In these experiments, a random-dot plane rotated about a vertical axis, and one or more patches of the plane protruded in depth toward the observer [5,32,33]. Animals performed better as the differences in depth among the patches increased, consistent with the animals perceiving depth from motion parallax. In addition, Schiller

2. Necessary factors for perception of depth from motion parallax

Is motion parallax a sufficient cue for depth perception? Although motion parallax had been appreciated by Helmholtz [16], it was not thought by some to be a valid cue to depth [17]. Many years passed before it was convincingly demonstrated that motion parallax is a sufficient cue for depth perception ([18], see [19] for a historical treatment). This was a large challenge because, in natural scenes, for humans and non-human primates, depth perception based on binocular disparity cues generally outperforms that based on motion parallax, both in precision [6,7] and accuracy [8,9]. For many species with laterally placed eyes, such as birds and rodents, the visual fields of the two eyes have much less binocular overlap than for primates. For these species, motion parallax cues to depth may be more important in many natural behaviours. For example, Ellard et al. [10] trained Mongolian gerbils to jump across a gap having variable distance. They observed that the frequency of head bobbing increased with the distance to be jumped. Since head bobbing generates motion parallax, these findings suggest that gerbils rely on motion parallax cues to estimate distance. Analogous results have been observed in other small animals [11,12].

Although it is clear that motion parallax cues make important contributions to three-dimensional vision, the neural basis of depth perception from motion parallax has been poorly understood, especially in comparison with the neural basis of stereopsis (see [13–15] for reviews). Here, we review important recent advances toward establishing a neural account of depth from motion parallax.

Figure 1. Similarity between motion parallax and binocular disparity as depth cues. (a) Motion parallax. If the head translates rightward, the image of a far object (open symbol, top) moves on the retina. If the eye moves through one inter-ocular distance, the position change on the retina due to motion parallax is equivalent to the object’s binocular disparity (as shown in panel b). Hence depth from motion parallax is often expressed in units of equivalent disparity. (b) Binocular disparity. Points falling along the geometric horopter, or Vieth-Muller circle (curved line), have zero binocular disparity. Here, the far object projects to disparate points in the retinal image for the two eyes (bottom). The binocular disparity shown here is equal to the change in position of the monocular image in a.
and colleagues reported that neurons in macaque primary visual cortex (V1) are selective to the relative depths of the patches [34].

While these studies are suggestive, they do not conclusively establish a neural substrate for perceiving depth from motion parallax. A limitation of the design of these experiments is that the ‘oddball’ stimulus could be identified simply by its retinal speed without necessarily relying on perception of depth. Therefore, speed selectivity is sufficient for V1 neurons to discriminate among stimuli having different depths in this physiological study [34]. Also, motion parallax in these stimuli was not coupled with any movements of the observer.

Indeed, to establish a neural correlate of depth from motion parallax, it is critical to dissociate depth from retinal image velocity, as achieved in psychophysical studies [18,23,25]. This led Nadler et al. [35] to design visual stimuli (figure 2c) that are depth-sign ambiguous when viewed by a stationary observer, but which yield clear percepts of depth-sign (near versus far) when viewed by a translating macaque that is trained to fixate a world-fixed target (viewing geometry illustrated in figure 2a,b). In this design, extra-retinal signals related to head or eye movements are necessary to disambiguate the depth-sign of the visual stimuli, thus providing a much more effective method to probe for a neural correlate of perceived depth.

With this stimulus paradigm, Nadler et al. [35] recorded extracellularly from neurons in the middle temporal (MT) area of macaque monkeys. Area MT is well known for its roles in visual motion perception [36], and also contains many neurons that are selective for depth defined by binocular disparity [37–39]. In addition, the MT area has been causally linked to stereoscopic depth perception [40–42], and therefore was a likely candidate to also signal depth from motion parallax.

Data from an example MT neuron are shown in figure 3a. In the retinal motion control condition, for which a stationary animal passively views the visual stimuli, the depth-tuning curve (blue) is symmetric around zero, such that near and far stimuli with the same depth magnitude elicit similar responses. In other words, responses of the MT neuron are depth-sign ambiguous in the absence of extra-retinal signals
related to head or eye movements. Responses increase with depth magnitude for neurons that prefer moderate to fast speeds (figure 3a), or decrease with depth magnitude for neurons that prefer slow speeds (figure 3b). Strikingly, when the animal was physically translated by a motion platform and counter-rotated his eyes to maintain fixation (motion parallax condition), the example neuron of figure 3a (black curve) becomes selective for depth-sign, with a clear preference for near depths. By contrast, other neurons prefer far depths in the motion parallax condition (e.g. figure 3b, black). Because the retinal image motion over the MT receptive field is the same in these two stimulus conditions (assuming accurate pursuit tracking, [35]), the difference in response between the retinal motion and motion parallax conditions must reflect the action of extra-retinal signals related to head translation or eye rotation. Note that the data shown by the green and orange curves in figure 3 will be described in a subsequent section.

Well over one-half of neurons in area MT showed significant selectivity for depth-sign in these initial experiments [35], thus establishing area MT as a potential neural substrate for computing depth from motion parallax. It is currently unknown whether similar selectivity appears in earlier stages of visual processing, such as V1, V2 and V3, that provide substantial inputs to MT [44,45].

4. Linking middle temporal activity to depth perception based on motion parallax

Although the findings of Nadler et al. [35] established that MT neurons carry information about depth-sign based on motion parallax, the experiments were performed in animals that were simply required to track a world-fixed target. To further understand the functional links between MT responses and depth perception from motion parallax, it is necessary to examine neural activity in animals that are trained to report their depth percepts. Kim et al. [6] trained animals to judge whether a random-dot stimulus, in which depth was only defined by motion parallax, appeared to be nearer or farther than the fixation target. The experimental approach was largely similar to that used by Nadler et al. [35], except that the proportion of dots that appeared at a near or far depth was variable and this ‘depth coherence’ parameter was used to titrate the difficulty of the task. Overall, Kim et al. [6] found that the sensitivity of individual MT neurons was lower than that of the animal, although the most sensitive MT neurons rivaled behavioural performance. This suggests that behavioural performance could be accounted for by pooling the responses of relatively small populations of MT neurons.

In addition, Kim et al. [6] examined whether the activity of MT neurons was predictive of perceptual decisions regarding depth by computing choice probabilities [46,47]. Choice probability quantifies the trial-by-trial correlation between responses of a single neuron and the animal’s choices, and generally reflects both correlated noise among neurons as well as the weights that are applied to neural responses in decoding [48]. The finding of significant choice probabilities is consistent with the possibility that neurons provide evidence used to perform the task, but it could also reflect top-down signals related to featureal attention or decision-making (see [47,49] for more extensive discussions of the interpretation of choice probabilities). In the depth-discrimination task based on motion parallax, Kim et al. [6] found that responses of many MT neurons were weakly predictive of the animals’ choices. The average choice probability for MT neurons was comparable with that found previously in studies of coarse motion discrimination [46,50], but smaller than that measured in a coarse depth-discrimination task based on binocular disparity cues [51]. The latter difference may stem from MT neurons having lower sensitivity to depth variations based on motion parallax cues than to depth variations based on binocular disparity cues [6].

Together, these findings from behaviors animals support the hypothesis that area MT provides important sensory information to inform perception of depth based on motion parallax cues. However, a causal test of area MT’s role in perceiving depth from motion parallax has not yet been performed, and such a test (e.g. based on electrical microstimulation or reversible inactivation, [40,42]) will be important to establish that MT is needed to perceive depth from motion parallax.

5. Sources of extra-retinal signals for depth-sign coding in area middle temporal

As described above, human psychophysical studies have demonstrated that eye movement signals, not head translation signals, disambiguate the sign of depth cued by motion parallax [24,25]. Thus, it was of considerable interest to establish whether the selectivity of MT neurons for depth-sign would show the same dependencies.

This issue was addressed by Nadler et al. [43], who extended the previous experimental approach [35] in two important ways. First, they introduced a head only condition in which the fixation target moved with the animal, such that no eye movements were necessary to maintain visual fixation (unlike in the scheme of figure 2a,b). In this head only condition, vestibular and other signals related to head/body translation were identical, but command signals to move the eyes were no longer needed. Second, they introduced an eye only condition in which the animal remained stationary but the virtual environment (including the fixation target) translated in front of the animal. This eliminated vestibular and other signals related to head/body translation, but required tracking eye movements. Findings from these conditions were compared with those obtained using the original design [35], in which both head and eye movements were involved (motion parallax condition). Importantly, visual stimuli over the receptive fields of MT neurons were the same across all these conditions.

The results reported by Nadler et al. [43] were remarkably clear. In the head only condition, MT neurons showed little or no depth-sign selectivity, and there was no correlation with results from the motion parallax condition (figure 4, open symbols). In striking contrast, results from the eye only condition were nearly identical to those from the motion parallax condition (figure 4, filled symbols). This indicates that the depth-sign selectivity of MT neurons depends on extra-retinal signals related to eye movements but not those related to head/body translation.

Whereas Nawrot [24] found reversed perception of depth-sign in humans tested with stimuli similar to our head only condition, we did not find reversed depth-sign tuning in MT responses during the head only condition. However, we were not able to maintain neural recordings
during head movements large enough to be likely to elicit a robust tVOR (see [43] for additional discussion). Thus, our findings are not in conflict with those of Nawrot [24].

The mechanisms by which eye movement command signals modulate responses of MT neurons to generate depth-sign selectivity require further study. However, preliminary findings suggest that eye movement signals powerfully modulate the response gain of MT neurons during specific directions of eye movements [52]. By analogy to the role of eye movement gain fields in performing coordinate transformations [53,54], a population of such gain-modulated MT neurons may be able to represent depth from motion parallax [52]. It should be noted that gain modulations of MT responses by pursuit eye movements have been reported previously [55], and it is possible that these effects could be related to the same mechanisms that generate depth-sign selectivity from motion parallax.

6. Role of visual cues in disambiguating depth-sign from motion parallax

Why should computation of depth from motion parallax, both behaviourally and in MT neurons, depend on eye movement signals and not signals (e.g. vestibular inputs) related to head translation? Important insights were provided by the theoretical work of Nawrot & Stroyan [27]. They demonstrated that depth from motion parallax can be computed from the following simple relationship, called the motion-pursuit law:

\[
\frac{d}{f} = \frac{d\theta}{d\alpha} \left( \frac{1}{1 - \frac{d\theta}{d\alpha}} \right) \approx \frac{d\theta}{d\alpha},
\]

where \( f \) is the fixation distance, \( d \) is the depth of an object relative to the fixation plane, \( d\alpha \) (which is shorthand for \( d\alpha/dt \))

denotes the change of eye orientation relative to the scene and \( d\theta \) (short for \( d\theta/dt \)) represents the retinal image motion of the point of interest (figure 5). Note that the approximation on the right holds when eye rotation is large relative to retinal motion. Critically, the motion-pursuit law requires knowledge of eye rotation relative to the scene, but does not involve head translation. Thus, this theoretical relationship helps explain why both perceptual and neural responses to motion parallax are disambiguated by eye movements but not head movements. Moreover, Nawrot et al. [8] have shown that psychophysical depth magnitude estimates are well described by a modified form of the motion-pursuit law.

The critical variable in the motion-pursuit law, \( d\alpha \), represents eye rotation relative to the scene. Thus, it is sensible that pursuit eye movement command signals, which control eye rotation relative to the head, would contribute to a neural representation of \( d\alpha \). However, any information related to \( d\alpha \) could be used, in principle, to compute depth from motion parallax. Could visual motion signals also be used to infer \( d\alpha \)? Optic flow is determined by translation and rotation of the eye relative to the scene, and theoretical work has shown that it is possible to decompose optic flow into rotation and translation components of observer motion [56,57]. Of specific interest here, whenever the eye rotates relative to the scene, the resulting images (under...
planar projection) contain systematic perspective distortions that appear as a component of ‘rocking’ motion in the image ([58], see supplementary movies 1–3). Therefore, under viewing conditions similar to those depicted in figure 2a,b, these ‘dynamic perspective’ cues could potentially be used to estimate the $\alpha$ term in the motion-pursuit law.

Early evidence that the visual system might process dynamic perspective cues to compute depth from motion parallax came from the work of Rogers & Rogers [59]. Observers viewed ambiguous depth corrugation stimuli while the visual display was rotated about a vertical axis through the centre of the display. This rotation of the display generates dynamic perspective cues similar to those resulting from eye rotations. Interestingly, Rogers & Rogers [59] found that this rotation of the entire display could disambiguate depth-sign in motion parallax displays. Eye movements were not monitored or enforced in these experiments, so it is difficult to be sure that the disambiguation resulted from dynamic perspective cues instead of systematic eye movements related to rotation of the display, but these results suggested a role for visual perspective cues in specifying $\alpha$.

Recently, Kim et al. [58] have provided a clear visual demonstration that dynamic perspective cues can disambiguate depth-sign (see their supplemental videos). Moreover, they have shown that MT neurons can unambiguously signal depth-sign based on motion parallax when dynamic perspective cues are present. Building upon the retinal motion stimulus condition used by Nadler et al. [35], they introduced a large stationary random-dot field around the small stimulus patch that was presented over the receptive field of an MT neuron. There was no head/body translation or eye rotation in this dynamic perspective condition, but rocking motion of the background dots simulated an eye rotation (figure 6a). Remarkably, many MT neurons showed robust depth-sign tuning in the dynamic perspective condition (e.g. figure 6b, magenta), even though the visual stimulus within the receptive field was identical to that of the retinal motion control condition for which MT neurons show no depth-sign selectivity (figure 6b, blue). Moreover, for many neurons, the depth-sign selectivity induced by dynamic perspective cues was similar to that induced by eye movements in the standard motion parallax condition (figure 6b, black).

It is worth noting that the visual stimuli for the retinal motion control condition also contain dynamic perspective cues [58]. However, when the stimulus is spatially restricted, these dynamic perspective cues are not sufficient to disambiguate depth-sign. This is consistent with recent psychophysical experiments showing that dynamic perspective cues in a small stimulus are not sufficient to disambiguate depth-sign [60]. Thus, dynamic perspective cues are mainly useful in large rich scenes, and thus it makes sense for the brain to make use of both dynamic perspective cues and eye movement signals to compute the critical variable ($\alpha$) needed in the motion-pursuit law.

These findings show that large-field patterns of image motion can be used to infer eye rotations relative to the scene, and that these signals can be used in useful neural computations instead of efference copies of eye movement command signals. In this light, it is worth noting that there is also evidence that dynamic perspective cues are used to
studies in humans [63–65] and monkeys [32] have provided evidence that depth perception involves integration of disparity and motion parallax cues, but until recently very little was known about neural integration of these cues. In humans, Ban et al. [66] used fMRI pattern analysis techniques to identify brain regions that appear to integrate disparity and motion parallax cues to depth. In animals, a few studies have examined how neurons signal three-dimensional surface orientation based on combinations of motion and disparity gradients [67] or perspective gradients and disparity gradients [68,69].

Recently, Nadler et al. [70] measured the depth-sign selectivity of macaque MT neurons based on both binocular disparity and motion parallax cues. One might expect neurons to prefer the same depth-sign (near or far) for each cue. Indeed, this was the case for many MT neurons, such as the example neuron illustrated in figure 7a. We refer to such neurons as ‘congruent’ cells. Surprisingly, however, other MT neurons prefer opposite depth-signs defined by disparity and motion parallax cues, and we refer to these as ‘opposite’ cells. Figure 7b shows an opposite cell that prefers near depths from motion parallax and far depths from binocular disparity. Overall, approximately 40% of MT neurons were found to prefer opposite depth-signs based on binocular disparity and motion parallax cues (data points in the top-left and bottom-right quadrants of figure 7c), whereas the remaining 60% have matched depth-sign preferences (top-right and bottom-left quadrants of figure 7c). We shall return to a potential functional role of opposite cells below.

Nadler et al. [70] also measured responses of MT neurons to stimuli involving congruent combinations of binocular disparity and motion parallax cues (combined condition). In general, responses in the combined condition reflected a mixture of selectivity for both depth cues. For the example congruent cell of figure 7a, depth-tuning in the combined condition (orange) appears to be dominated by binocular disparity tuning (green), but tuning strength is slightly enhanced in the combined condition, reflecting the influence

![Figure 7](http://rstb.royalsocietypublishing.org/)
of motion parallax inputs. For other neurons, such as the opposite cell illustrated in figure 7b, depth-tuning in the combined condition (orange) appears to be dominated by motion parallax tuning (black), and selectivity in the combined condition is slightly reduced by the incongruent disparity selectivity. These findings suggest that congruent cells might contribute to perceptual cue integration of depth cues, whereas opposite cells would not, but the study of Nadler et al. [70] was not designed to test these issues directly. Additional studies, in which both psychophysical and neuronal performance is tested with both congruent and conflicting combinations of disparity and motion parallax cues, will be needed to evaluate whether activity of MT neurons can account for perceptual cue integration and cue weighting (as explored previously for multisensory neurons involved in heading perception, [71–73]).

Thus far, we have focused on perception of depth of objects that are stationary in the world. However, it is often important to detect objects that move in the world. When an observer translates through the environment, all objects in a scene (except those on which the eye fixates) will move in the retinal image, which greatly complicates the task of detecting objects that move in the world. Conflicts between binocular disparity and motion parallax cues may play important roles in solving this problem, as illustrated in figure 8. As the observer translates to the right, objects that are nearer than the plane of fixation move to the left, whereas farther objects move to the right. The speed of image motion depends on the distance of an object from the fixation target. Thus, if the depth of an object is known (e.g. from binocular disparity cues), then the expected image velocity could be computed under the assumption that the object is stationary in the scene. For an object that moves in the world (such as the car in figure 8), the image velocity will deviate from this prediction. Thus, for an independently moving object, binocular disparity and motion parallax cues to depth would be in conflict, and it is possible that the brain could detect a moving object by identifying this cue conflict (without having to perform calculations that depend on self-motion velocity, viewing distance, etc.). Does the brain have neural mechanisms that can detect this cue conflict to identify objects that are moving in the world?

Human psychophysical studies suggest that the visual system is sensitive to conflict between disparity and motion parallax cues. Early evidence came from Rogers & Collett [74], who examined how humans perceive conflict between disparity and motion parallax cues in random-dot stimuli. Interestingly, such conflicts led subjects to perceive rotation of the stimulus relative to the scene. More recently, Rushton et al. [75] showed that human observers can efficiently detect an independently moving object among stationary distractors when there is an appropriate conflict between motion parallax and binocular disparity cues. This suggested that the brain may contain specialized mechanisms for detecting objects that move in the world during observer motion.

Nadler et al. [70] have proposed that opposite cells in area MT, which have opposite depth-sign preferences for motion parallax and binocular disparity cues, may play an important role in detecting moving objects during self-motion. Specifically, opposite cells may have greater activity than congruent cells when stimulated by an object that moves in the world, because the retinal image motion for such an object is not consistent with its binocular disparity. A preliminary report by Kim et al. [76] provides further support for this idea. Animals were trained to perform a task similar to that used by Rushton et al. [75], in which they detect an object that moves in the world relative to other objects that are stationary in the scene. Interestingly, in a subset of conditions in which all objects had identical retinal motion, MT neurons that responded more strongly to objects that move in the world were found to predict trial-by-trial variations in the animals’ choices [76]. This supports the notion that the brain may use a mixture of congruent and opposite cells to detect moving objects during observer translation.

8. Summary and future directions

Our understanding of how the brain computes depth from motion parallax has evolved substantially over the past decade, and this review has summarized some of the major advances. These advances provide an excellent example of how psychophysical, theoretical and neurophysiological studies can inspire each other to achieve a deeper understanding of an advanced perceptual function. At the same time, our understanding of this problem is still quite limited and there are a number of fruitful directions for additional research.

Thus far, depth-sign selectivity has only been observed in area MT, and we do not know whether the computations required to generate depth-sign selectivity are initiated in MT or whether MT inherits this selectivity from its inputs. Additional experiments in areas that provide direct input to MT, including areas V1, V2 and V3, will be valuable for establishing the circuitry underlying computations of depth from
motion parallax. Along these same lines, the neural origins of the pursuit command signals and dynamic perspective signals that disambiguate depth-sign in MT neurons, as well as the detailed mechanisms of their operation, remain unclear.

The critical variable (d) in the motion-pursuit law [27] is a change of eye orientation relative to the scene. In the experiments conducted so far, eye orientation relative to the head has been varied. More generally, as we move and fixate on a target in the world, eye rotation relative to the scene may be determined by some combination of eye rotation relative to the head, head rotation relative to the body, and body rotation relative to the scene. Thus, it is possible that multiple extra-retinal signals are used to disambiguate depth. Future experiments should examine whether perception of depth from motion parallax depends on a combination of these rotation variables or just eye rotations relative to the head. Corresponding neural studies could examine whether neurons in area MT (or elsewhere) also receive additional signals (e.g. vestibular rotation inputs) that can disambiguate depth-sign. In addition, it will be important to examine whether and how the brain combines different sources of signals related to d. For example, are extra-retinal inputs related to eye, head and body rotations combined to provide a more general representation of d that is used to perceive depth? In this regard, it is worth noting that dynamic perspective cues provide direct information about eye rotation relative to the scene. Thus, a representation of d based on multiple extra-retinal signals might be integrated with dynamic perspective cues to provide a unified, multimodal d signal that will be robust across many environmental conditions.

Finally, depth from motion parallax has generally been studied during lateral translation of observers. Although this simplifies the geometry and has a number of advantages for experimental design, our movements through the environment often have substantial (if not predominant) fore–aft components. Fore–aft translations are studied extensively in the domain of heading perception [77–79]. Psychophysical studies have shown that the human visual system can ‘parse out’ objects from optic flow patterns [80,81], but we know little about how humans or animals estimate depth of objects under these conditions. Thus, there may be an opportunity to integrate domains of research related to motion parallax and heading perception.

Together, these observations suggest that studying the neural computations that mediate perception of depth from motion parallax will continue to provide exciting opportunities to understand how visual and extra-retinal signals are combined in specific ways to generate specialized neural representations that underlie behaviour.

Authors’ contributions. H.R.K. conceptualized the review, wrote first draft, edited drafts and approved final version; D.E.A. conceptualized the review, edited drafts and approved final version; G.C.D. conceptualized the review, wrote first draft, edited drafts and approved final version.

Competing interests. We have no competing interests.

Funding. H.R.K. and G.C.D. were supported by a grant from the National Eye Institute (EY013644). This work was supported by NIH grant no. EY013644 (to G.C.D.) and by an NEI CORE grant no. (EY001319).

References


