Vision Research 234 (2025) 108653

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

The contributions of pictorial, motion, and binocular cues to the perception of depth and distance

Paul B. Hibbard *, Jordi M. Asher, Rebecca L. Hornsey

Division of Psychology, University of Stirling, Stirling, FK9 4LA, Scotland, United Kingdom
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, Essex, England, United Kingdom

ARTICLE INFO

Keywords:
Binocular vision
Motion parallax
Perspective
Pictorial cues
Depth estimation
Cue combination
Natural images

ABSTRACT

Multiple visual cues are available for the estimation of distance. According to the modified weak fusion model, the information from these cues is combined through weighted averaging, with the weights determined by the relative reliability of each cue. Empirical tests of this model tend to isolate a small number of cues, in order for their reliabilities to be manipulated. Weights measured in this way are specific to the testing environment, and do not allow us to quantify the contributions of individual cues in natural viewing. To address this, we used estimates from the literature of sensitivity for a wide range of distance cues to predict the contribution of pictorial, binocular, and motion cues to relative distance. The cues assessed included convergence, accommodation, height in the field, texture density, relative size, height in the field, binocular disparity, and motion (assuming a walking observer). We used the modified weak fusion model to estimate the contribution of binocular, motion, and pictorial cues for distances between 2 and 100 m. These calculations provide estimates of the expected contributions of individual depth cues in everyday viewing conditions. In most cases, our results show a clear benefit for the weighted averaging of cues in the natural environment, in comparison with the use of the most reliable cue alone.

1. Introduction

Our visual experience of the three-dimensional world is created from the two-dimensional images formed in the retinas of our two eyes. These retinal images contain a wealth of cues to the position and shape of three-dimensional objects and surfaces (Cutting & Vishton, 1995). These cues are typically classified based on the source of the information that they provide. *Pictorial* cues are those that are available from a single retinal image at a given moment in time and are what allow us to see the three-dimensional layout in pictures. They include cues such as perspective, shading, occlusion, the height and size of objects and features in the image, texture gradients, and aerial perspective. *Binocular* cues are those that result from having two eyes. They can be classified as retinal (binocular disparity: the differences between the two retinal images) and extra-retinal (binocular convergence: the difference in the direction of gaze of the two eyes). *Motion* cues are those that depend on changes in sensory information over time. Again, these cues can be classified as retinal (the changes in the retinal images over time) or extra-retinal (the movements of our eyes, head, and body).

Depth cues vary in the kinds of information that they provide about the three-dimensional location of objects relative to the observer. Here, we use the term *distance* to refer to the distance between a point in space and the observer, and *depth* to refer to the difference in distance between two points. To emphasise that depth is measure of difference, we refer to distance values as d and depth values as Δd. Some cues, such as binocular vision and motion parallax, provide information about the absolute distance to objects and the depth between points. Others, such as occlusion, merely provide information on which of the two points is closer. The existence of multiple cues is important for reliable depth perception because it creates redundancy in the available information. This provides the two important benefits of robustness and precision.

Given that not all potential depth cues may be available in all scenes, as long as one of the several cues is available the observer will be able to estimate depth. For example, texture cues will not be present when viewing an untextured surface, binocular cues when viewing with one eye, or motion cues when the observer and all objects in the scene are stationary. However, as long as at least one of the many potential cues is available, the observer will be able to estimate the depth. The availability of multiple cues thus provides robustness against the absence of any particular cue.

Access to multiple cues also increases the precision with which depth can be estimated. Due to the imprecision of sensory information,

* Corresponding author at: Division of Psychology, University of Stirling, Stirling, FK9 4LA, Scotland, United Kingdom.
E-mail address: paul.hibbard@stir.ac.uk (P.B. Hibbard).

https://doi.org/10.1016/j.visres.2025.108653
Received 15 May 2023; Received in revised form 16 February 2025; Accepted 6 June 2025
Available online 2 July 2025
0042-6989/

each cue does not provide a single point estimate for each location in a scene. Rather, a likelihood function can be estimated, expressed as the relative probability of the sensory information s_i as a function of depth Δd:

$$p(s_i|\Delta d) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(\Delta d - \Delta d_i)^2}{2\sigma_i^2}} \tag{1}$$

Here, the likelihood function is modelled as a Gaussian distribution defined as a function of depth Δd. s_i is the relevant sensory information, which may for example be the difference in binocular disparity between two points. It is typically assumed that this is unbiased (so that Δd_i is equal to the actual depth between the targets) and the imprecision of the estimate is determined by the standard deviation σ_i. It defines the relative probability that the (known) sensory information s_i was generated by each possible value of Δd. The reliability of the cue, $r_i(d)$, is defined as the reciprocal of the variance $\sigma^2(d)_i$:

$$r_i(d) = \frac{1}{\sigma^2(d)_i} \tag{2}$$

Where variance, and therefore reliability, will vary as a function of distance d. When multiple cues are available, Bayesian inference provides a framework for calculating an optimum decision given likelihood functions from multiple cues, and prior information about the structure of the world (Landy, Maloney, Johnston, & Young, 1995; Tyler, 2020). With independent Gaussian likelihood functions, reliability can be optimised by taking a weighted average of all the available cues:

$$\Delta d_s = \sum_{i=1}^{n} w_i \Delta d_i \tag{3}$$

where the weights are given by:

$$w_i = \frac{r_i}{\sum_{j=1}^{n} r_j} \tag{4}$$

When cues are combined in this way, and assuming unbiased and independent estimates with Gaussian uncertainty, the precision of our depth estimates will increase and will always be greater than the precision that could be achieved from one cue, or a subset of the available cues (Cochran, 1937; Landy et al., 1995). Since this weighted-averaging approach is equivalent to multiplying the individual-cue likelihood functions, then taking the peak of the resulting distribution, it is also known as maximum likelihood estimation (MLE) (Rohde, van Dam, & Ernst, 2016). Note that this procedure involves the use of all available cues, including any that may be very unreliable, but that the weight assigned to each cue is proportional to its reliability. Provided that these weights are set correctly according to the relative reliabilities of the cues, this weighted averaging will increase the reliability of the resulting estimate:

$$r_s = \sum_{i=1}^{n} r_i \tag{5}$$

This is illustrated in Fig. 1a, for a simple case that contains only two cues. For each individual cue, the likelihood is an unbiased normal distribution. The likelihood function that results from combining these cues is also unbiased, but has a smaller standard deviation, and thus a greater reliability. When the means of the two single-cue likelihood functions differ, the mean of the combined-cue likelihood function will be a weighted average of the two (Fig. 1b).

Psychophysical experiments have demonstrated these predicted improvements in the precision of perceptual estimates. While there are many studies in this area (Trommershäuser, Körding, & Landy, 2011), we illustrate a number of the important attributes of cue combination through three examples.

The first is a study of the combination of visual and haptic cues to object size (Ernst & Banks, 2002). Participants judged the size of a horizontal bar that they viewed binocularly and/or grasped with their thumb and index finger. When only one cue was available, the just noticeable difference (JND) in bar height was smaller for the visual than the haptic condition. When both cues were available, and the reliability of the visual cue was reduced by adding noise to the stimulus, the JND was smaller in the combined-cue condition than in either of the single-cue conditions. This study established that cues can be combined to increase the reliability of estimates. The relative weights assigned to the two cues by the participants in the study were also well predicted by the maximum-likelihood cue-combination model.

The second example is a study of the combination of two visual cues – binocular disparity and texture – in the perception of the slant of planar surfaces (Hillis, Watt, Landy, & Banks, 2004). JNDs for slant defined by each cue were measured as a function of the slant of the surface, and its distance from the observer. The reliability of disparity as a cue to slant decreased with distance and varied with slant in a manner that depended on distance. For surfaces defined by texture, reliability increased with increasing distance and was unaffected by slant. When surfaces were defined by both cues, reliability increased in a manner predicted by the maximum likelihood cue-combination model. These results demonstrated that the changes in relative reliability of the two cues with changes in slant and distance were taken into account in weighting their estimates. Also, when the slants defined by the two cues differed, perceived slant was close to the predicted weighted average of these two values.

The third example also evaluated the combination of binocular disparity and texture cues, but in this case assessed how the cues are combined to provide an estimate of the size of an object that is used to control grasping (Keefe, Hibbard, & Watt, 2011). This is a particularly salient assessment of whether people are able to access all available information in performing a task, since it has been argued in contrast that reaching and grasping may be specialised to rely preferentially on binocular cues (Goodale & Milner, 2013; Marotta, Behrmann, & Goodale, 1997; Melmoth, Finlay, Morgan, & Grant, 2009; Melmoth & Grant, 2006; Watt & Bradshaw, 2000). At near distances, the reliability of binocular cues was at its highest and no improvement was observed from the addition of texture cues. However, the reliability of binocular cues reduced with increasing distance to a greater extent than the reliability of texture cues. A clear improvement in precision as a result of cue combination was observed at the furthest distance tested, at which the reliabilities of the two cues were equally matched. These results highlight how the reliabilities of individual cues and the benefits of cue-combination can vary with properties such as the distance to objects. They also show how these properties can be manipulated in an experimental design in order to provide a strong test of the cue-combination model. In this case, testing at a relatively far distance ensured that the reliabilities of the two cues were sufficiently well-matched to show that reaching and grasping are not specialised for the use of binocular cues.

Studies such as these provide empirical tests of the maximum likelihood cue-combination model, since this model provides clear predictions of how the stimulus property under investigation will be perceived, and the accuracy with which it can be estimated. These predictions are particularly powerful because they vary with stimulus manipulations that form the core of the designs of the experiments.

What this approach cannot establish however, is the contribution that these cues will make to the reliability of depth perception in everyday vision. To see this, it is useful to compare the situation in a typical laboratory experiment with everyday experience. In the laboratory, researchers go to great lengths to remove, or control, cues that are not the focus of the experiment. The study by Hillis et al. (2004) is an exemplary case. Participants were positioned using a bite-bar in order to remove motion cues, and a diffuser was positioned in front of the screen to blur the image, thus reducing the influence of accommodation, blur gradients and the phosphor grid of the screen. In the binocular condition, sparse random dot stereograms were used to remove pictorial depth cues such as texture density and linear perspective. The texture cues were complementary, being presented monocularly to remove all binocular depth cues. The rendering of the stimuli also meant that shading cues (Todd, 2020) were absent. All of

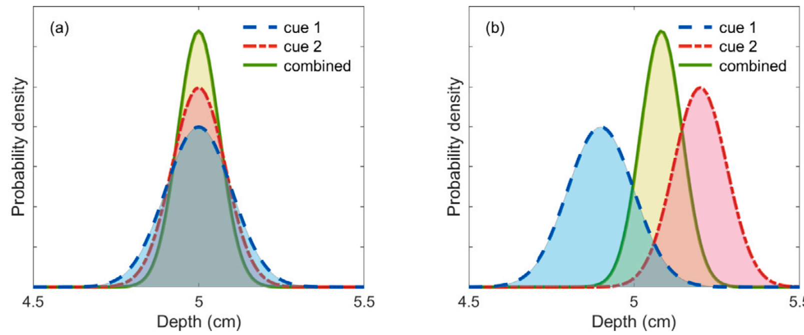

Fig. 1. (a) Unbiased, Gaussian likelihood functions for two depth cues. The likelihood function of the combined-cue estimate has a smaller standard deviation. (b) If the two cues are biased, with different means, the mean of the combined-cue estimate will lie between the two.

these measures were used to carefully isolate the two cues of interest, and minimise all other sources of information. The weights measured in these experiments thus represent the relative weights of the cues when all other information is minimised. They cannot be used in isolation to predict the weights attributed in the natural environment when the full gamut of cues is available.

The second important feature of the design of these experiments is that the relative reliabilities, and therefore the predicted weights, of the cues are manipulated within the experiment as part of the design. This was achieved by adding noise (Ernst & Banks, 2002), or by exploiting the natural variation of cues with distance and slant (Hillis et al., 2004; Keefe et al., 2011). Although these manipulations demonstrate the power of studies in testing the mechanisms of cue combination, the predicted relative weights are set by these experimental conditions, rather than reflecting the contributions that might be found in a typical everyday setting.

Koenderink (1998) argued that, rather than studying vision under reduced cue conditions, we should study it at an operating point in which all cues are available and kept constant, and only one cue is manipulated. This then allows us to assess the importance of the manipulated cue in everyday viewing. This approach was adopted by McCann, Hayhoe, and Geisler (2018) to assess the degree to which binocular cues improve depth sensitivity. Participants were presented with stereoscopic views of natural scenes, for which ground truth depth information was available. Depth sensitivity for determining which of two points was closer was measured as a function of distance in the scene, and lateral separation in the image. Sensitivity to depth differences was proportional to absolute distance, and observers were able to make very precise judgements, with Weber fractions as low as 0.3%. Sensitivity was better than that modelled from just considering the cues of binocular disparity and image size, consistent with observers making good use of the rich variety of depth information available in natural scenes. The contribution of binocular cues was estimated by comparing performance under monocular and binocular viewing. The increased sensitivity under binocular viewing is consistent with typical binocular disparity thresholds, which they estimated to be 25 arc sec.

This approach differs from the typical cue-isolation methodology (Palmisano, Gillam, Govan, Allison, & Harris, 2010), which is used to demonstrate how well observers can judge depth from individual cues, but does not assess the importance of this sensitivity in natural viewing conditions. This is an important limitation since, even if the reliability of a cue is known, for example from measurements in a reduced-cue environment, it is only possible to predict its weight in determining precision in more typical viewing conditions if the reliabilities of all other available cues are known.

The purpose of the current study is to provide a better understanding of how a range of cues contribute to the perception of depth in the natural environment. First, we estimate the relative contributions of pictorial, binocular, and motion cues to the perception of depth in everyday scenes. Second, we quantify how these contributions increase the reliability of depth estimates. Finally, we demonstrate how these contributions vary with factors including the distance between the observer and the targets, and the movement of the observer. We performed our calculations for three distance ranges: near (personal) space, intermediate (action) space, and far (vista) space (Cutting & Vishton, 1995).

Here, near space is defined as the space within 2 m of a person, and is the region within which we are able to interact with objects. Intermediate space is defined as the range from 2 to 30 m and it is within this space that we plan our upcoming movement and actions and undertake many social interactions. Far space is defined as distances beyond 30 m. In addition to the enhanced effectiveness of parallax depth cues at near distances, close objects are larger in the image than far away ones and are therefore easier to discriminate and more salient (Sousa, Brenner, & Smeets, 2011; Sousa, Smeets, & Brenner, 2012). There is evidence of the human visual system prioritising stimuli presented in near space compared to far (Kaas & Mier, 2006; Làdavas, 2002), which has been suggested to be an evolutionary response for reflexes in response to harmful stimuli and defencive behaviour (Blini et al., 2018; Graziano & Cooke, 2006; Makin, Holmes, Brozzoli, & Farnè, 2012; Sambo, Forster, Williams, & Iannetti, 2012). Due to the differences found between performance in tasks when stimuli are presented within different areas of space, three categories of distance were tested in the current experiment. Irrespective of the theoretical status of this categorical division of space (Previc, 1998), it is helpful to assess the way in which the contributions of cues vary with distance. This is because, since the reliability of parallax cues reduces rapidly as distances increases, the weights associated with all cues will be expected differ across these distance ranges. Motion cues were excluded in the near-space analysis, since these are modelled based on the assumption of a moving (walking) observer. The analysis in this case therefore represents a simulation of a stationary observer, for example at work at a desktop or workbench.

In addition to evaluating the value of depth cues in the natural environment according to the maximum likelihood approach, we also considered how their contribution would be affected by a different choice of decision rule. Two closely related alternative rules that have been considered are (1) the minimum sigma rule and (2) probabilistic cue switching (Jones, 2016; Scarfe, 2022). Under the minimum switching rule, judgements are based only on the most reliable cue available. The improvement in performance against that predicted by the best single cue is taken as the hallmark of maximum likelihood estimation, and evidence that the fusion of sensory information occurs (Rohde et al., 2016). Under the probabilistic cue switching rule, again only one cue is used on any given judgement, with the more reliable cues being chosen more often (Byrne & Henriques, 2013; De Winkel, Katliar, Diers, & Bülthoff, 2018; Nardini, Jones, Bedford, & Braddick, 2008; Serwe, Drewing, & Trommershäuser, 2009). Both of these rules provide the key benefit of robustness that results from the availability of multiple cues. However, the reliability of the resulting estimate will be less than that provided by maximum likelihood estimation with appropriately chosen weights. The reliability of probabilistic cue switching will also be less than that of minimum sigma, since judgements will sometimes be based on one of the less reliable cues that are available. These rules do not however require the fusion of individual cues into an overall all-cues estimator. We thus compared the predictions of maximum likelihood and minimum sigma estimation to quantify the benefit of this fusion process.

2. Materials and methods

Cutting and Vishton (1995) modelled how JNDs in the distance between two points vary with distance, for a wide variety of cues. In all cases, an assumption was made that the points to be compared are nearby in the image (separated by less than 5 degrees of visual angle), and viewed foveally. We used these estimates to predict the contributions of these cues. JNDs were expressed by Cutting and Vishton as depth contrast, the difference between the near and far points, divided by the mean of the two:

$$C = \frac{2(D_2 - D_1)}{D_1 + D_2} \tag{6}$$

However, for our analysis, we are interested in the difference in depth only ($\Delta d = D_2 - D_1$), rather than the depth contrast.

2.1. Parallax distance cues

A number of the cues (height in the field, binocular disparity, convergence, motion parallax, and accommodation) can be described as parallax cues, from which distance can be estimated through triangulation. For a parallax cue, a baseline of known length is defined, and depth can then be calculated from the difference in direction of the target from the points at each end of this baseline. For the cues listed above, the baselines are defined by the height of the observer's eyes above a plane, their interocular distance, the movement of their head from one time to another, or the width of their pupil, respectively. The geometry of each of the parallax cues is shown in Fig. 2.

The baseline B is defined by the person's height H for height-in-the-field, pupil size P for accommodation, interocular separation I for binocular vergence and disparity, and the distance moved between timepoints T_1 and T_2 for motion parallax.

If the length of the baseline is B, the distance to the nearer of the two points is D_1, and the smallest change in visual angle that can be detected is δ, then the JND $\Delta d = D_2 - D_1$ occurs when the further point is at a distance:

$$D_2 = B \tan\left[\left(\arctan \frac{D_1}{B}\right) + \delta\right] \tag{7}$$

In the case of height-in-the field, the relationship shown in Eq. (7) can be related directly to the right-angled triangle formed by height H and distance D in Fig. 2. For binocular fixation on an object directly in front of the observer, each eye's direction forms a right angle with the cyclopean direction, giving

$$D_2 = \frac{I}{2} \tan\left[\left(\arctan \frac{D_1}{\frac{I}{2}}\right) + \frac{\delta}{2}\right] \tag{8}$$

where $delta = \alpha_1 - \alpha_2$. With a small angle approximation this can be simplified to

$$D_2 = I \tan\left[\left(\arctan \frac{D_1}{I}\right) + \delta\right] \tag{9}$$

2.1.1. Height in the field

This cue applies to objects which rest on a plane, typically the ground plane for a standing observer. If we assume that this plane is horizontal and that the height of the observer's eyes above the plane is known, then the distance to an object is monotonically related to the angle of declination from straight-ahead when viewing the object (Fig. 2a). When looking straight ahead, this equates to the vertical position in the retinal image, with objects further away appearing higher up in the image. For a standing observer with objects on the ground plane, the baseline is determined by the observer's height. We used a height of 1.7 m, the current average height of an adult in the United Kingdom (Moody et al., 2013). Following Cutting and Vishton (1995), we assumed a JND for height in the field of $\delta_{HIF} = 5$ arc min.

The same geometry also applies to other situations, such as a seated observer interacting with objects on a table-top, or a standing observer reaching for objects on a kitchen work surface. In these cases it is the height of the observer's eyes above the surface, rather than the ground plane, that is important. In our calculations for near space, we assume an eye-height of 50 cm above a table-top. For comparison, eye-height above a kitchen work surface would be around 75 cm, with this increased baseline providing more reliable distance information.

2.1.2. Accommodation

Light from a point in space will diverge across a range of directions before arriving at the pupil. The eye's optics must converge these rays to a single point to create a sharp image on the retina (Fig. 2b). Thus accommodation and the degree of blur of a point while focussing on a location at a different distance can also be used as parallax cues to the relative depth between two points (Held, Cooper, & Banks, 2012). Our estimate of the reliability of accommodation used an assumed pupil diameter of 5 mm and a JND of $\delta_{accommodation} = 50$ arc sec (Nagata, 1991).

2.1.3. Binocular convergence (vergence)

When we fixate an object binocularly, the two eyes are oriented in slightly different directions. If we assume that the observer's interocular distance is known (Taya, 2023), then this binocular convergence angle (or vergence) can be used to estimate distance (Fig. 2c). We assume an interocular distance of 64 mm (Dodgson, 2004) and a JND for a difference in vergence of $\delta_{vergence} = 10$ arc min Nagata (1991).

2.1.4. Motion parallax

When we move relative to an object, we sample the ambient optic array from multiple locations. The visual information provided by comparing the sampled array from two such locations is geometrically equivalent to that provided by binocular cues (Fig. 2d). When we move horizontally through a distance equivalent to our interocular separation, our sensitivity to depth from motion parallax is similar to that for binocular disparity (Rogers & Graham, 1982). Cutting and Vishton (1995) modelled the sensitivity to motion parallax to be slightly higher than that for binocular disparity; in our model, we matched the level of sensitivity they proposed using a baseline of 64 mm and a JND of $\delta_{motion} = 20$ arc sec.

2.1.5. Binocular disparity

When we fixate on an object, its location on the retinal image is the same for both eyes. Other points in the scene will appear at different locations in the two eyes, depending on their distance relative to the fixated point, creating binocular disparities. These disparities are crossed for objects nearer than fixation and uncrossed for objects farther than fixation (Fig. 2e). Binocular viewing can thus provide information about the distance to many points in the scene, not just the one that is fixated. We again assume an interocular distance of 64 mm (Dodgson, 2004), and a JND for a difference in distance from binocular disparity of $\delta_{disparity} = 25$ arc sec Nagata (1991). Note that, although based on the same underlying geometry, the sensitivity to relative differences in distance from the retinal cue of binocular disparity is much higher (here by a factor of 24) that for the extra-retinal cue of vergence.

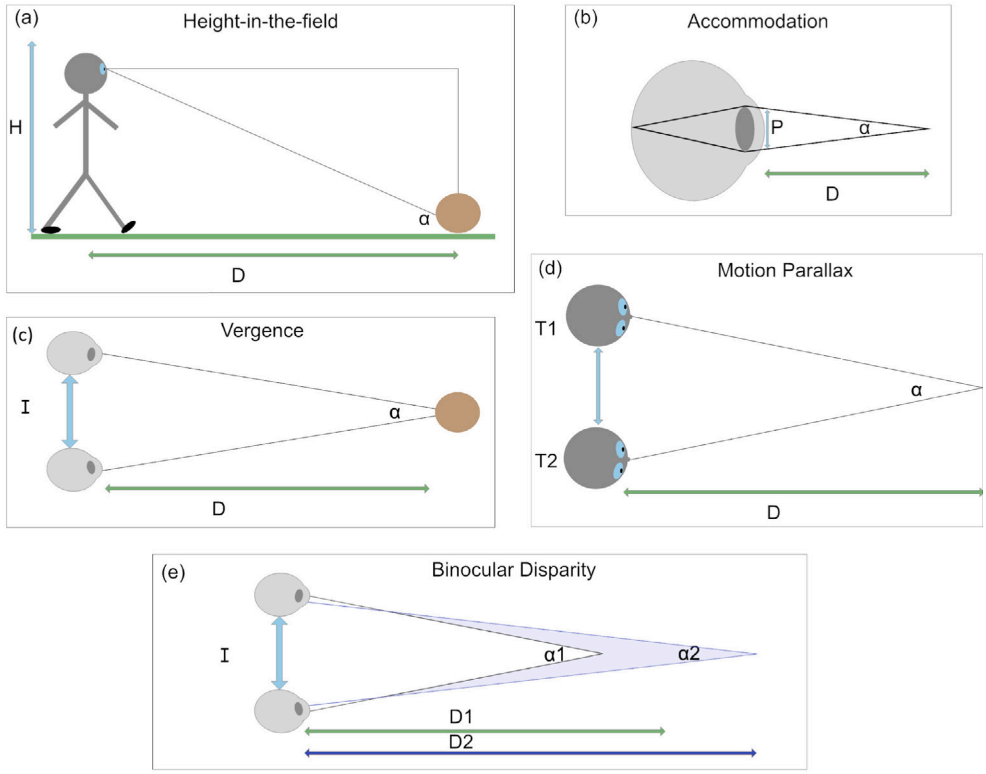

Fig. 2. Parallax distance cues allow distance to be estimated by triangulation, given a baseline of known length. This baseline is defined by (a) the height of the observer for height-in-the-field; (b) the width of the pupil for accommodation; (c) the interocular distance for vergence (d) the movement of the observer for motion parallax; and (e) the interocular distance for binocular disparity.

2.2. Non-parallax distance cues

2.2.1. Relative size and density

The projected size of an object will decrease with its distance from the observer. This means that, for objects or texture elements that are on average equally sized and equally densely distributed in space, relative image size and density can be used to judge relative distance (Aloimonos, 1988; Gibson, 1950). That is, for a constant size and density of features on a surface, the size of image features will increase, and their density decrease, as distance is reduced. Comparison of these properties around two spatial regions can be used as a cue to their relative distance. Cutting and Vishton (1995) assumed that observers could detect a difference of 3% in size and 10% in density, regardless of the absolute difference. These values were used to estimate JNDs that increased linearly with absolute distance.

2.2.2. Occlusion

Most objects in our environment are opaque, such that when one object is closer than another, and positioned so as to block the line of sight to parts of the further object, it will occlude the view of the latter. This means that occlusion provides a simple, reliable cue to which of two objects is closer. However, it is limited in only providing information about the depth order of the two objects, rather than information about the size of the depth difference. It is also only available when the objects are positioned laterally so that one occludes the other. Nevertheless, when it is present, occlusion provides a very reliable depth cue. We assume that observers can detect a depth difference of 0.1% from this cue (Cutting & Vishton, 1995). We provide calculations separately for cases of occluded and unoccluded objects.

2.3. Calculation of single-cue reliability

JNDs were taken to be the difference in distance at which observers would be able to correctly judge which of two points was closer 75% of the time in a two-alternative forced choice procedure. These values were taken from the estimates presented by Cutting and Vishton (1995). The standard deviation of the uncertainty in each cue is then given by $\frac{1}{\sqrt{2}}$ of this value (Green & Swets, 1974). For each cue, reliability was calculated from Eq. (2).

2.4. Calculation of cue weights, reliability and JNDs for multi-cue stimuli

For each combination of cues, the optimal weight for each cue was calculated from Eq. (4), and the reliability of the resulting estimator from Eq. (5). The JND of the multi-cue estimator was then calculated from this estimate of reliability.

3. Results

3.1. Near space

3.1.1. Cue reliability

The reliabilities of the cues are plotted in Fig. 3 as a function of distance. These are plotted in three ways, as (1) depth contrast (Fig. 3a); (2) JNDs (Fig. 3b); and (3) reliabilities (Fig. 3c). Depth contrast values were used to calculate JNDs — the smallest depth (difference in distance) that could be discriminated with 75% accuracy from each cue. These in turn are also plotted as the cue reliability, the inverse of the estimated standard deviation of the likelihood function for each cue. For all cues, reliability decreases with distance. The most reliable cues are occlusion, binocular disparity, and height in the field.

3.1.2. Cue weights

The reliability measures plotted in Fig. 3 determine the weights that should be attributed to each of the available cues for a given estimation. These weights are plotted in Fig. 4. Calculations are provided for estimates with and without an occlusion cue to the relative depth of

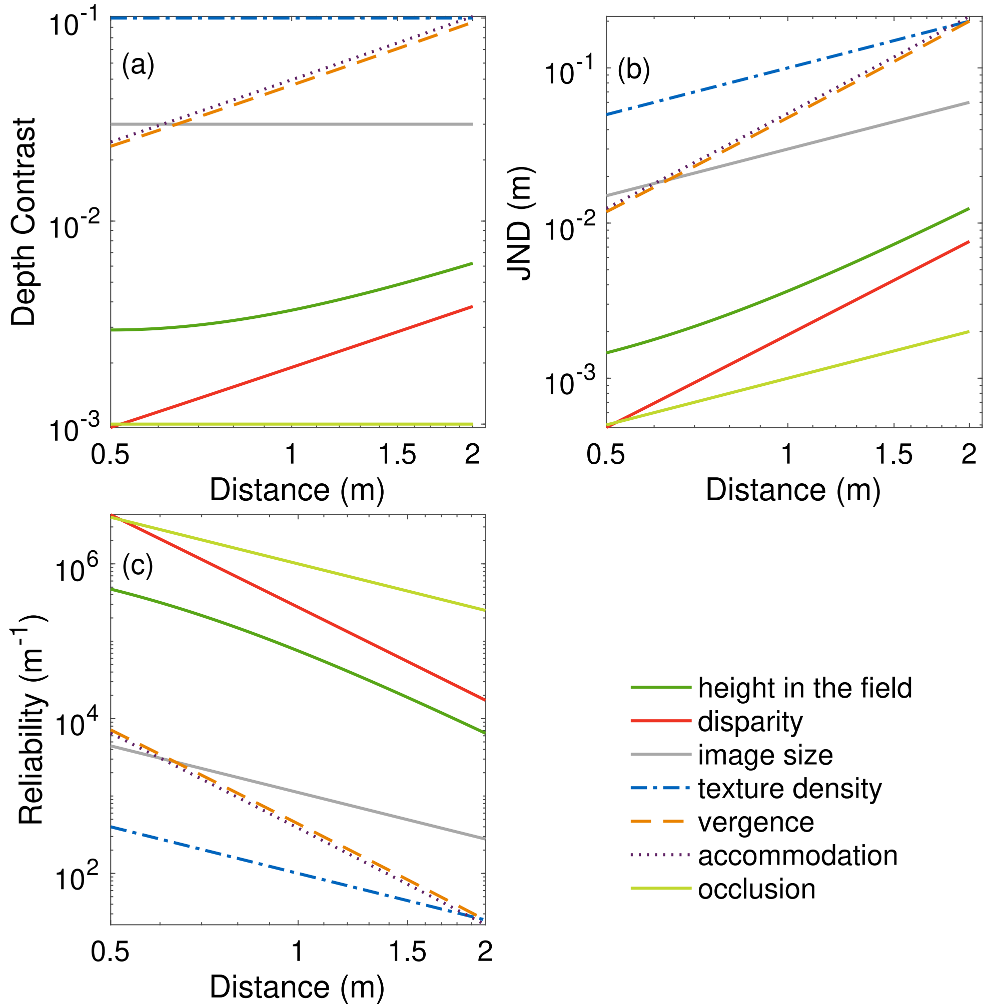

Fig. 3. (a) Depth contrast as a function of distance in near space. This is calculated as the JND in distance between two targets, divided by the average of the two distances. (b) JNDs as a function of distance. (c) Reliabilities for each cue, calculated as the inverse of the standard deviation of the likelihood function.

target points for a static observer. Occlusion is manipulated because it is a very reliable, and therefore a dominant cue when it is present, but also one that is not always available.

When occlusion is available, its weight increases from 31% at 50 cm to 47% at 2 m. The weight for binocular disparity reduces from 49% at 50 cm to just 6% at 2 m. Together, these two cues thus contribute between 53 and 81% of the information for depth judgements in this range. When occlusion is available, the other binocular parallax cue of vergence is predicted to make no practical contribution (< 0.1 %) to ordinal depth judgements.

Depth sensitivity for unoccluding points is dominated by binocular disparity, with a weight varying between 91% at 59 cm and 56% at 2 m. The second most reliable cue in this case is height in the field, with a weight varying from 8% at 50 cm to 42% at 2 m. Across the near distance range, the combined weight of these cues is estimated to be 99% or above, meaning that no other cues would make a meaningful contribution to this task when these cues are available.

In near space, depth cues are important for the control of action, such as reaching to and grasping objects. It has been suggested that binocular cues play an especially important role in this task (Goodale & Milner, 2013; Marotta et al., 1997; Melmoth et al., 2009; Melmoth & Grant, 2006; Watt & Bradshaw, 2000). The combined weights for the binocular cues of disparity and vergence, and the improvement in JND that these provide, are plotted in Fig. 5. When judging the relative depth between two unoccluding points, the weight attributed to binocular cues is calculated to exceed 90% at the nearest distance, leading to a 70% reduction in the JND. Thus, even without a specialisation for the use of binocular cues, they will tend to dominate the estimation of depth within reaching distance.

3.1.3. *MLE versus MS decision making*

JNDs for the MLE and MS decision rules were compared for monocular and binocular viewing, and occluding and unoccluding pairs of points (Fig. 6). In this analysis, there were no motion cues but all other cues were available. In all cases, the JNDs for the MLE rule were smaller than those for the MS rule, with this difference increasing with distance. However, the improvement was only substantial when binocular cues were available. These results demonstrate that there is a benefit to cue fusion in near space.

3.2. *Intermediate space*

3.2.1. *Cue availability*

There is a practical limit on the farthest distance at which parallax cues can be used to make relative depth judgements. As distance increases, visual directions from the observer to each target will tend towards parallel. Given the uncertainty of sensory estimates, this limit occurs when the closer point creates an angle of convergence that is one JND, and the farther point is at some distance beyond this. If both points are beyond this limit, the difference in the convergence angles to the two will be less than one JND.

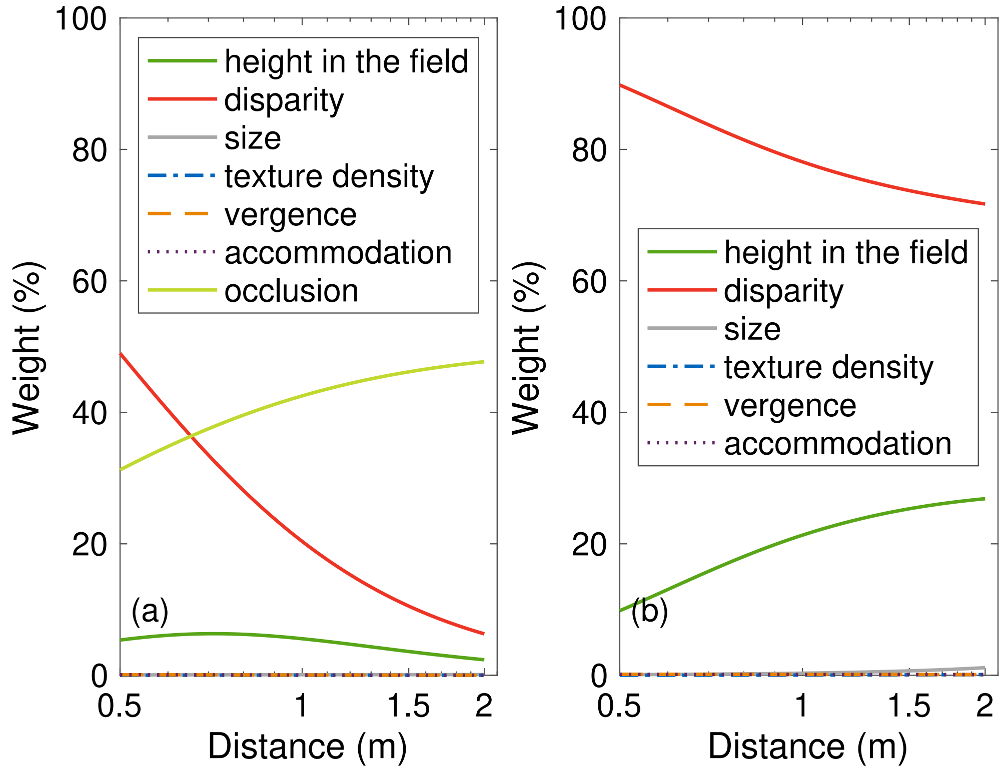

Fig. 4. Depth cues weights as a function of distance, for (a) the case in which occlusion is available; and (b) when occlusion is absent. Binocular disparity and occlusion are the dominant cues.

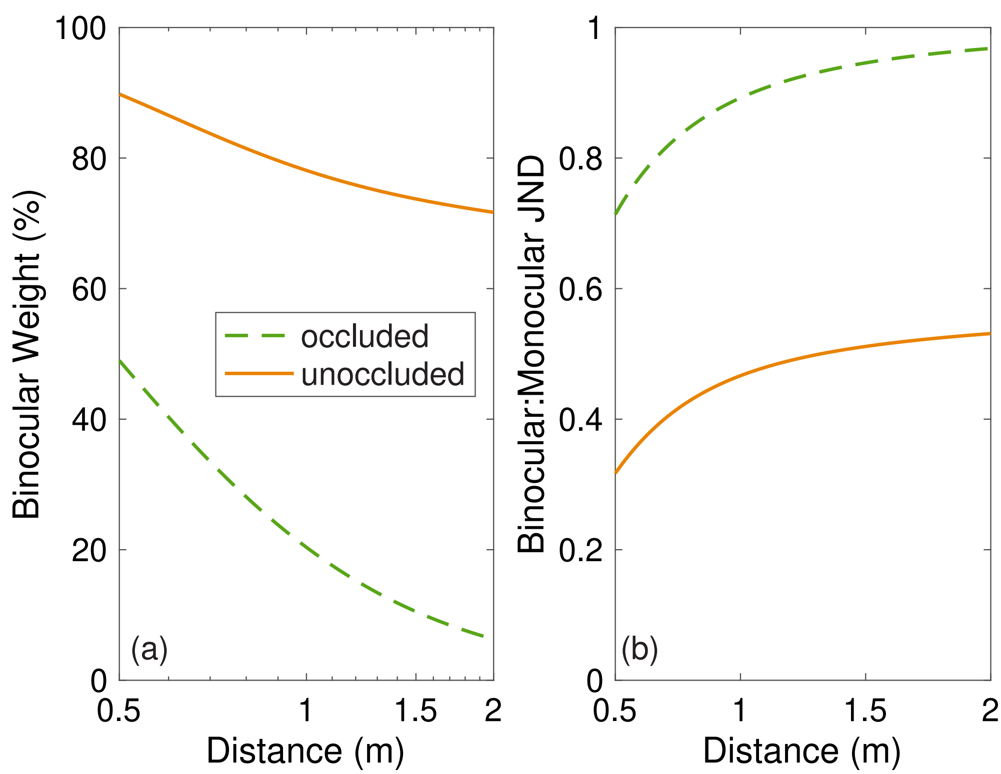

Fig. 5. (a) The combined weight of the two binocular cues of disparity and vergence. These are plotted separately for occluded and unoccluded objects. When occlusion is not available, the binocular weight rises above 90% for the nearest distance. (b) The reduction in JND that results from the presence of binocular cues, calculated as the ratio of JNDs for the case where all cues are included, and the case where binocular disparity and vergence are excluded. Binocular cues contribute to a reduction of JND of 70% at the nearest distance.

This limit is reached at 22 m for vergence and 20.6 m for accommodation. The other parallax cues (height in the field, binocular disparity, and motion parallax) will be available in principle across the whole of the intermediate distance range.

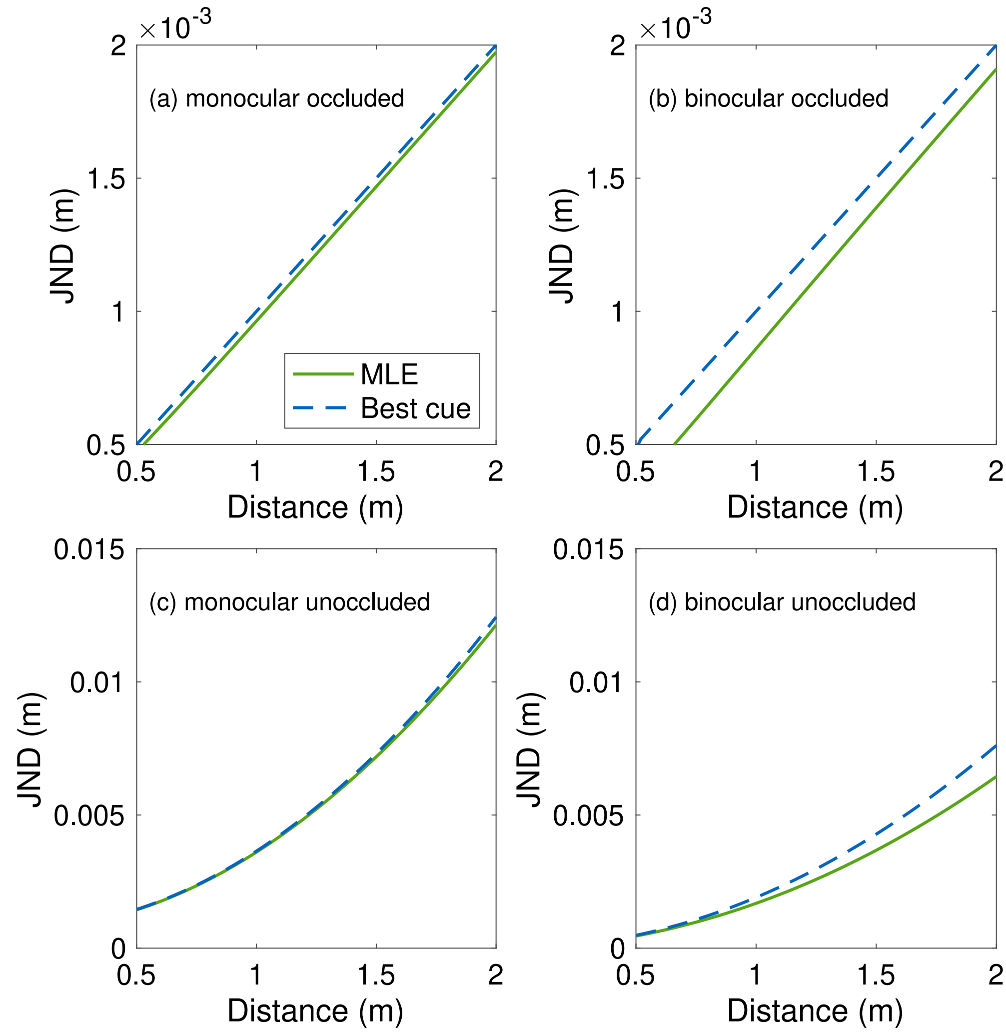

Fig. 6. JNDs in near space for the maximum-likelihood and best-cue decision rules. The decrease in JND for the MLE rule quantifies the degree of improvement in precision provided by cue fusion. Data are plotted as a function of distance for (a) monocular occluded; (b) binocular occluded; (c) monocular unoccluded; and (d) binocular unoccluded points.

3.2.2. *Cue reliability*

Depth contrasts, JNDs and cue reliabilities are plotted for intermediate space (2 to 30 m) in Fig. 7. For all cues, reliability decreases with distance. The most reliable cues are occlusion, height in the field, and binocular disparity.

3.2.3. *Cue weights*

Cue weights for the intermediate distance range are plotted in Figs. 8 and 9. They are plotted for four illustrative conditions, since the weight for each cue is determined not only by its reliability, but also by which other cues are present. Calculations are provided for scenes with and without an occlusion cue to the relative depth of target points, and for a static and dynamic observer. Motion is manipulated since it is a cue that is dependent on the observer's behaviour, and not determined only by the scene itself. In intermediate space, whether an observer is static or walking will affect the weights of all other cues.

Both Fig. 8a and 8b show that, when available, occlusion will again be a dominant cue to relative depth. This is true for both static and moving observers. For a static observer, the weight for occlusion varies from 46% at 2 m to 50% at 30 m. For a moving observer, the weight is very similar, increasing from 44% at 2 m to 50% at 30 m. When occlusion is not present, the parallax cues of height-in-the-field, motion parallax and binocular disparity are more heavily weighted (Fig. 9a and 9b). Of these three, height-in-the-field is the most reliable, reflecting the much larger baseline (the observer's height) associated with this cue.

3.2.4. *Binocular cues*

In Fig. 10 the expected contributions of binocular cues to depth judgements are shown. Estimates are provided with and without occlusion, and for a moving and static observer, and it is assumed that all other cues are available. The binocular weight is calculated as the sum of the weights for the two binocular cues of vergence and disparity. As expected from geometrical considerations, in all cases the binocular weight decreases with distance. When a usable occlusion cue is present, the importance of binocular cues is relatively modest, with a maximum weight of 3%. When occlusion is not available, the maximum binocular weight increases to 37% for a static observer, and 24% for a moving observer.

We also calculated the expected improvement in depth sensitivity provided by binocular vision, expressed as the ratio of JNDs under monocular and binocular viewing (Fig. 11a). When occlusion is available, the expected improvement is very small, not exceeding 3%. Without occlusion however, the estimated improvements are considerable, reaching 21% for a static observer, and 13% for a moving observer. It is notable that binocular cues make a substantial contribution at the farthest distances in this range. For example, for a static

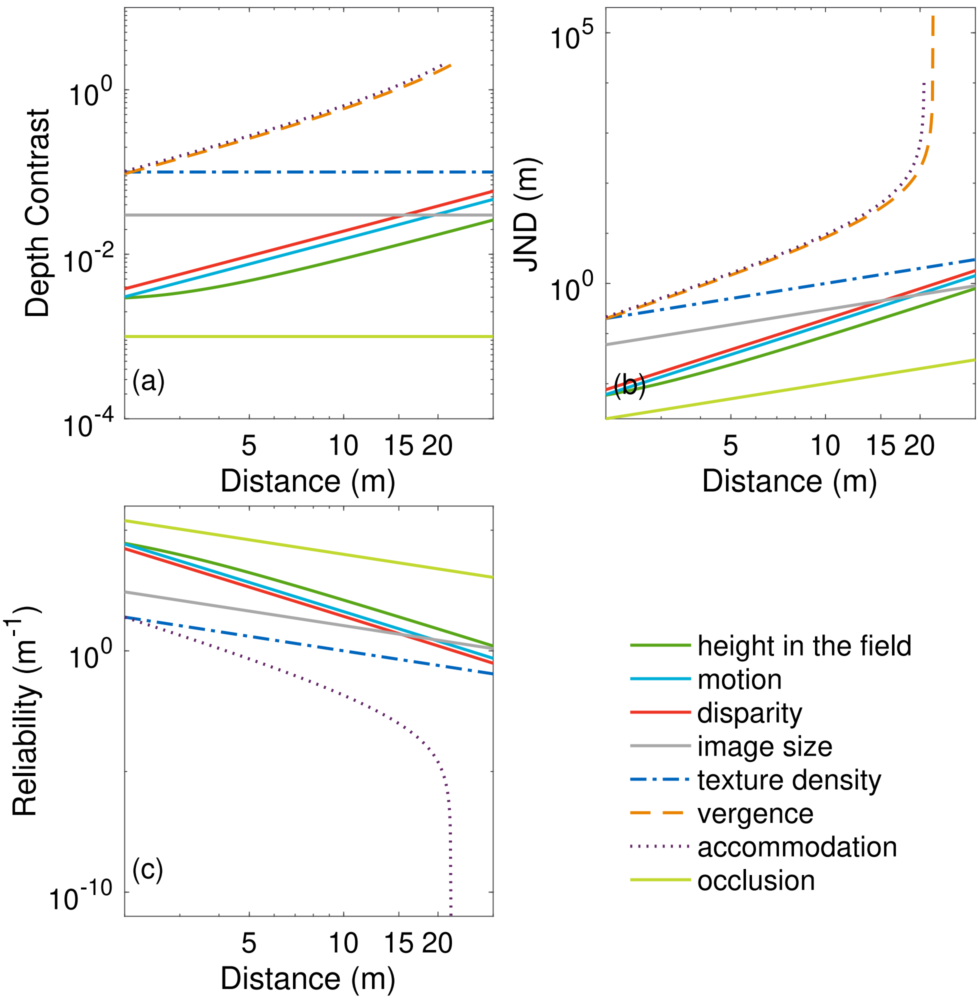

Fig. 7. (a) Depth contrasts (b) JNDs (c) reliabilities of cues in intermediate space, plotted in the same way as in Fig. 3.

observer, without occlusion, a reduction in JND of 4.8% is expected at a distance as far as 30 m.

3.2.5. Motion cues

The improvement in JND provided by motion parallax, expressed as the ratio of JNDs for a static and moving observer, is plotted in Fig. 11b. When occlusion is available, the expected improvement is again small, not exceeding 5%. Without occlusion however, the maximum estimated improvement in JND is 2% under binocular viewing, and 28% under monocular viewing.

3.2.6. MLE versus MS decision making

JNDs for the MLE and MS decision rules were compared for monocular and binocular viewing, and occluding and unoccluding pairs of points, for both a static (Fig. 12) and a moving (Fig. 13) observer. All other cues were available in all cases. For unoccluding pairs of points, JNDs for the MLE rule were smaller than those for the MS rule, for both a static and a moving observer, with this difference increasing with distance. These results demonstrate that there is a benefit to cue fusion in intermediate space. For occluding pairs of points, the superior reliability of occlusion in this distance range means that cue fusion provides no improvement in precision.

3.3. Far space

3.3.1. Cue availability

Accommodation and vergence are not included since the maximum distance for these cues falls within intermediate space. All other cues were calculated as being available up to at least the farthest distance of 100 m considered. For the remaining parallax cues, the maximum distances were calculated as 528 m for binocular disparity, 660 m for motion parallax, and 1.1 km for height in the field.

3.3.2. Cue reliability

Depth contrasts, JNDs and cue reliabilities for far space (30 to 100 m) are plotted in Fig. 14. Across this range occlusion, height in the field, and image size are the most reliable cues.

3.3.3. Cue weights

Cue weights for the far distance range are plotted in Fig. 15 for occluding points, and Fig. 16 for unoccluding points. Weights were calculated for a static and moving observer. For occluding points, weights for occlusion were greater than 99%, meaning that no other cues would contribute to relative depth judgements in these cases. Weights are thus only plotted for unoccluding pairs of points.

For both moving and static observers, the two most important cues are image size and height in the field. The former increases in weight with distance, while the later decreases. For a moving observer, the weight for motion parallax decreases from 13% at 30 m to 2% at 100 m.

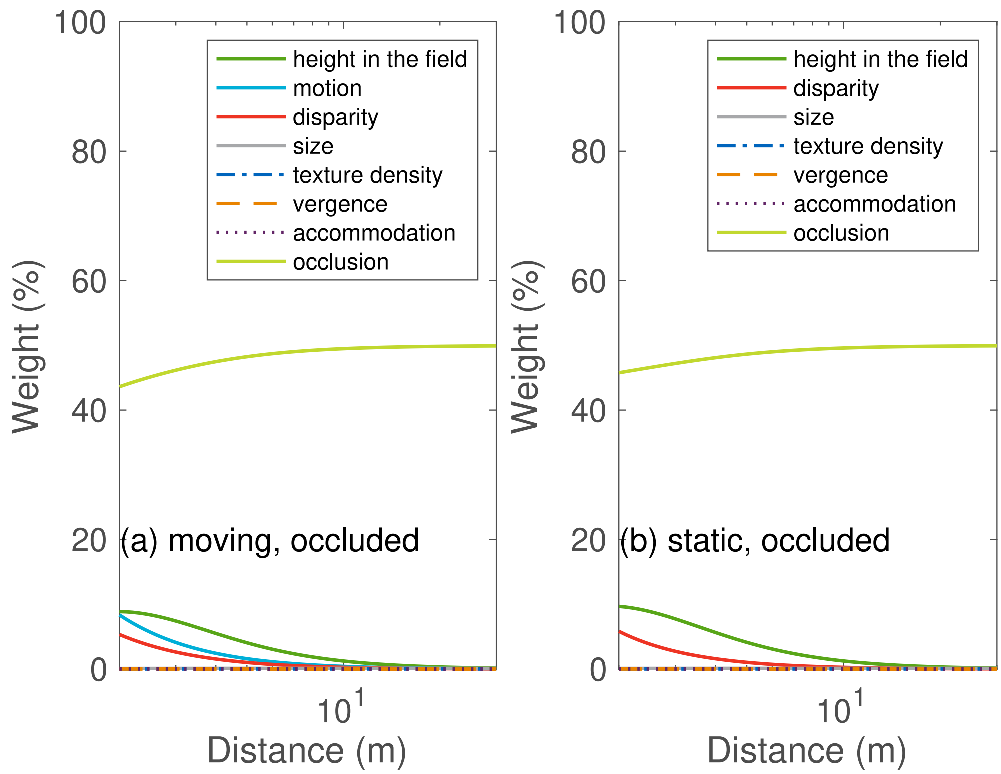

Fig. 8. Cue weights for the intermediate range for (a) occluding points for a moving observer; (b) occluding points for a static observer. Results are similar in both cases, with occlusion providing the most reliable cue.

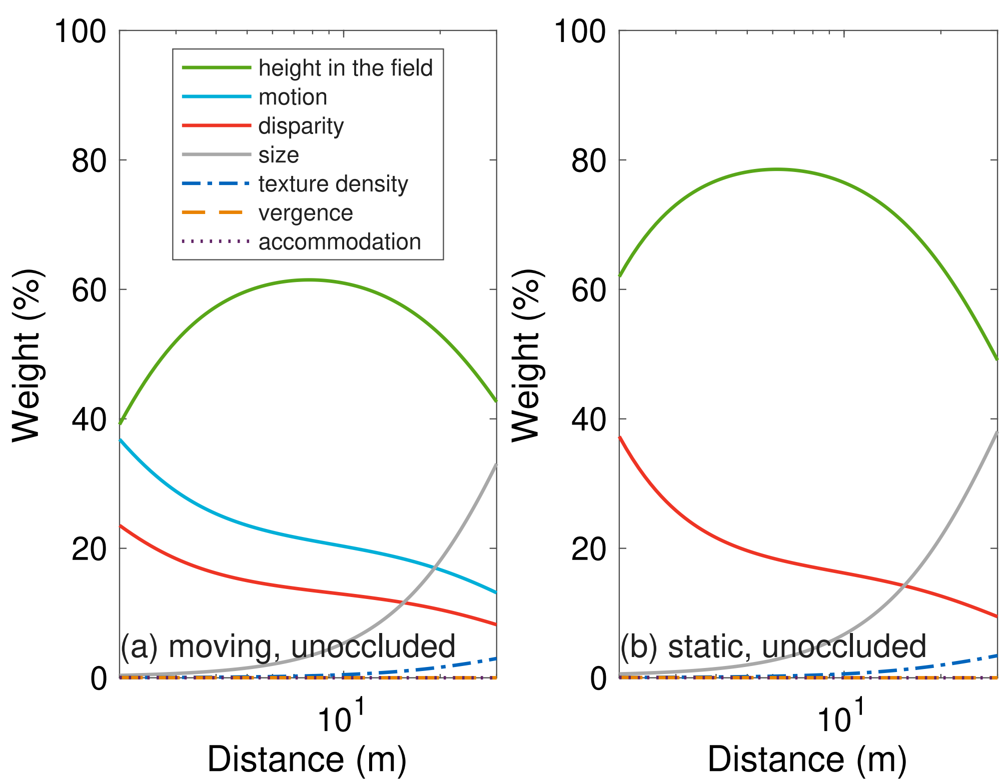

Fig. 9. Cue weights for the intermediate range for (a) unoccluding points for a moving observer; and (b) unoccluding points for a static observer.

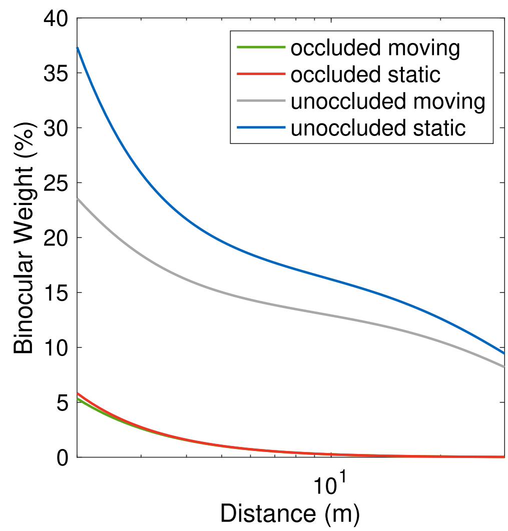

Fig. 10. The combined weight of the binocular cues of disparity and vergence. Data are plotted for a moving and a static observer, and in the presence or absence of occlusion.

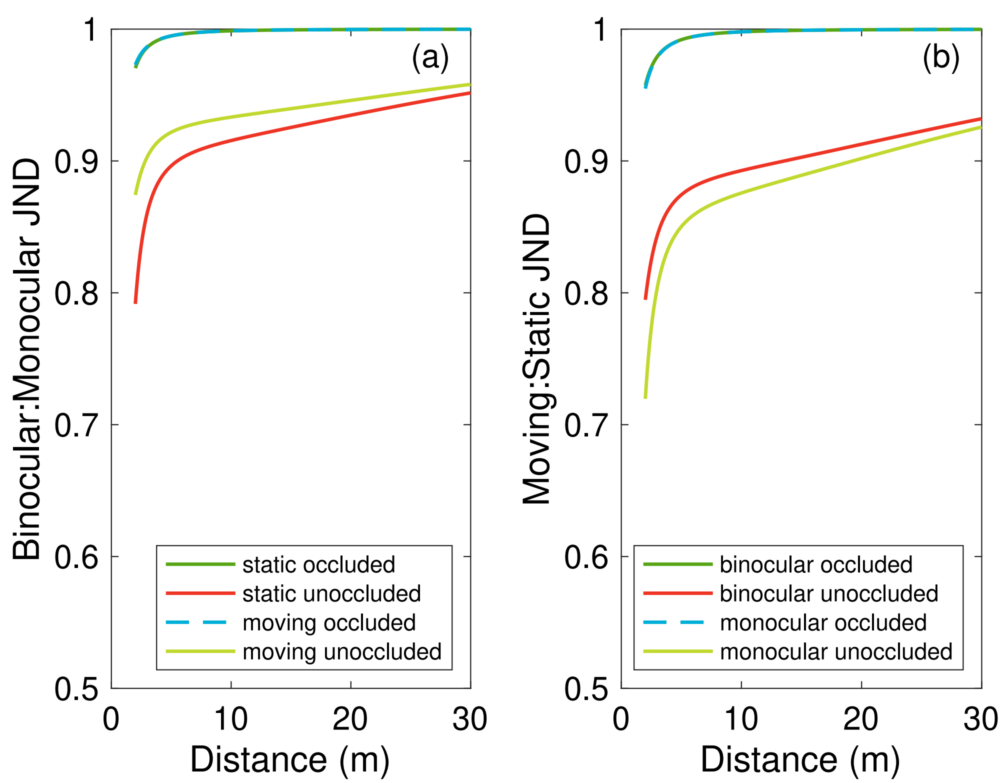

Fig. 11. The reduction in JND provided by (a) binocular cues; and (b) motion parallax in intermediate space. JNDs were calculated as function of distance, taking all cues into account and excluding binocular or motion cues. The improvement in JND is plotted as the ratio of the binocular and monocular (or moving and static) JNDs.

3.3.4. Binocular cues

Improvements in JNDs provided by binocular vision are plotted in Fig. 17a. With occluding points, binocular cues make no appreciable contribution to depth sensitivity within this distance range. For non-occluding points, improvements fall from around 5% at 30 m, for both static and moving observers, to less than 1% at 100 m.

3.3.5. Motion cues

Improvements in JNDs provided by motion parallax are plotted in Fig. 17b. With occluding points, motion cues also make no appreciable contribution to depth sensitivity within this distance range. For non-occluding points, improvements fall from around 7% at 30 m, to around 1% at 100 m.

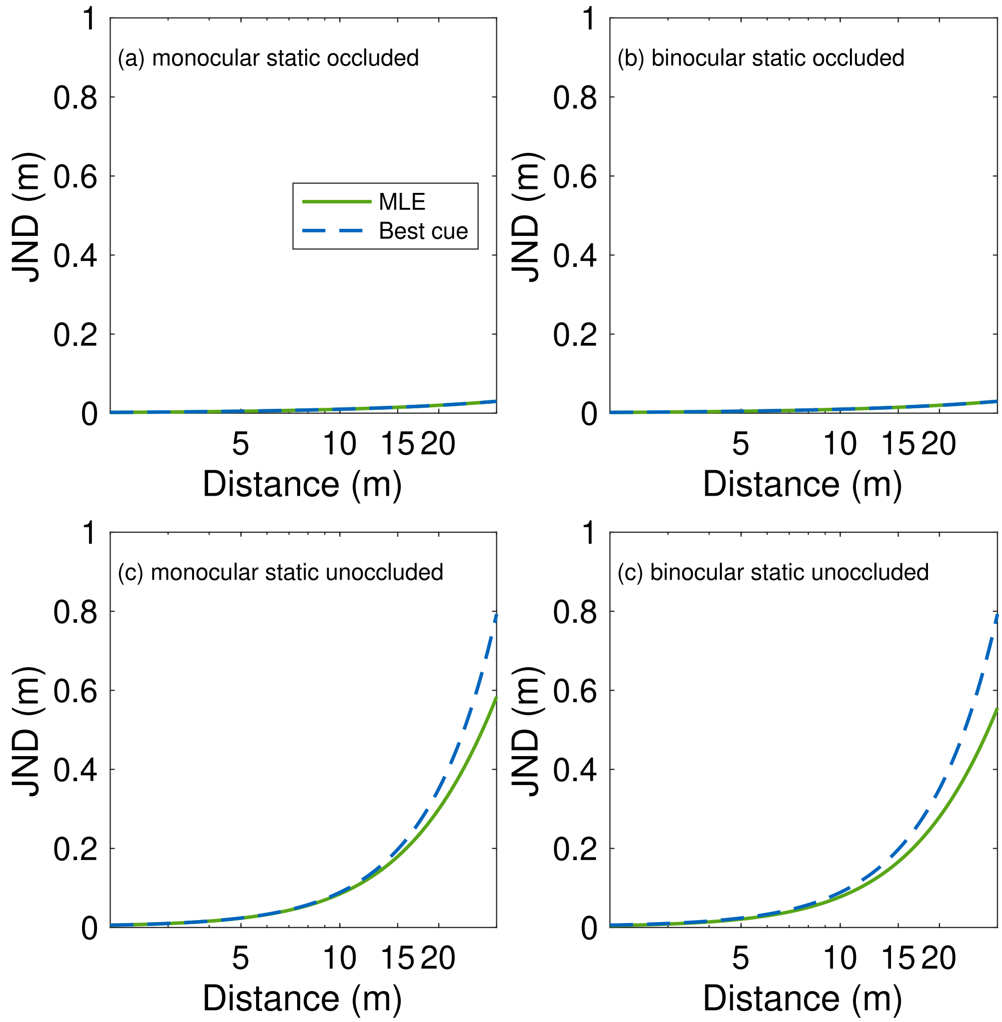

Fig. 12. JNDs in intermediate space based on a Maximum Likelihood model (MLE) and a decision rule in which depth is estimated by using only the most reliable cue, with the smallest standard deviation (MS), for a static observer. (a) Monocular viewing with occlusion; (b) Binocular viewing with occlusion; (c) Monocular viewing with no occlusion; (d) Binocular viewing with occlusion. A clear benefit is observed when binocular cues are available.

3.3.6. *MLE versus MS decision making*

JNDs for the MLE and MS decision rules were compared for monocular and binocular viewing, and occluding and unoccluding pairs of points, for both a static (Fig. 18) and a moving (Fig. 19) observer. For unoccluding pairs of points, JNDs for the MLE rule were smaller than those for the MS rule, for both a static and a moving observer, with this difference increasing with distance. These results demonstrate that there continues to be a benefit to cue fusion in far space. However for occluding pairs of points, the superior reliability of occlusion in this distance range means that cue fusion provides no improvement in precision.

3.4. *Simulating individual differences*

Individual differences in factors such as height, inter-pupillary distance (IPD), and pupil diameter could influence depth cue reliability. While a full exploration of these variations is beyond the scope of this paper, we ran the analysis for an observer with an IPD or height that is two standard deviations above or below the mean, covering approximately 95% of the population. The standard deviations used were 3.8 mm for IPD, and 6.5 cm for height. The effects of IPD on the binocular weight, and the improvement in JND provided by binocular cues, are shown in Figs. 20 and 21, for near and intermediate space respectively. This analysis illustrates the expected range of variation in cue weights and improvements in reliability. These data show the modest effects of IPD variation in the adult population. The effect for far distance, where the contributions of binocular cues is very small, is shown in Fig. 22.

The effects of changes in observer height in the weighting of the height-in-the-field cue are shown in Fig. 23. Results are shown for intermediate and far space only, since in these cases a standing observer was modelled. Again, only relatively modest effects on the weights are predicted, and the effects of distance are unchanged.

4. Discussion

The purpose of this study was to estimate the contributions made by a broad range of pictorial, binocular, and motion cues to the estimation of relative distance in typical everyday environments. Our findings are summarised for each class of cue below.

4.1. *Cue contributions as a function of distance range*

4.1.1. *Occlusion*

Occlusion is a very potent cue that tends to dominate other depth cues when it is available. In the near distance range we estimate that its contribution is between 31% and 47%. This increases with distance, to between 46% and 50% in the intermediate range, and over 99% in the far range. It is therefore a very dominant cue, with this dominance increasing with distance. This reflects the fact that it is uniformly

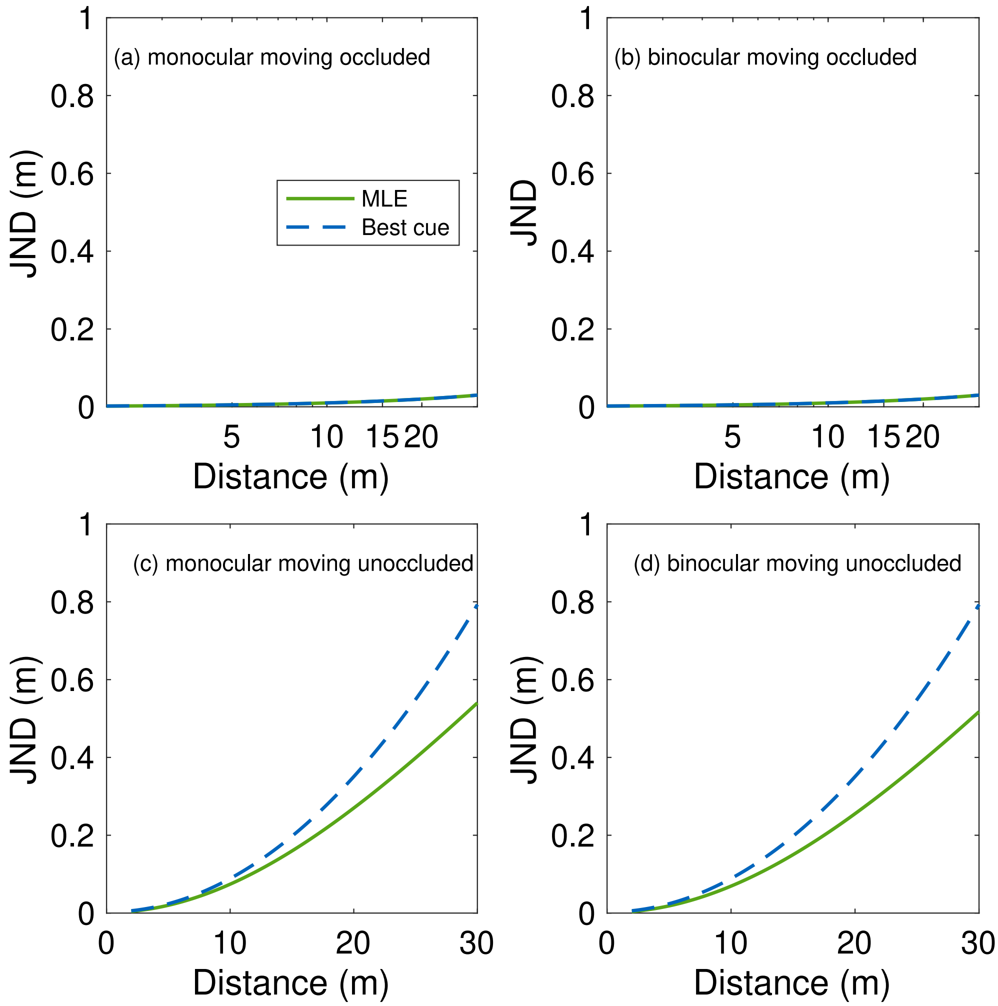

Fig. 13. JNDs in intermediate space based on a Maximum Likelihood model (MLE) and a decision rule in which depth is estimated by using only the most reliable cue, with the smallest standard deviation (MS), for a moving observer. (a) Monocular viewing with occlusion (b) Binocular viewing with occlusion (c) Monocular viewing with no occlusion (d) Binocular viewing with occlusion. A clear benefit is observed when binocular cues are available.

reliable, in contrast with parallax cues, whose reliability decreases rapidly with distance. Occlusion is limited however in providing only ordinal depth information about which of two points is closer, and in only being available when those points are on surfaces that overlap.

Despite these geometrical limitations, analysis of natural images has shown that, when combined with information about the convexity of concavity of occluding contours, occlusion is able to provide a degree of metric depth information (Burge, Fowlkes, & Banks, 2010).

4.1.2. Binocular cues

The binocular cues of disparity and vergence make a substantial contribution to the reliability of depth estimates for nearby points, with this contribution diminishing with distance. For non-occluded objects, the weight assigned to binocular cues reduces from between 71% and 90% in near space, to 6% in intermediate space, and less than 1% in far space. When occlusion is available, these weights reduce to between 6% and 49% in near space. In intermediate space, the maximum weight is 37% for a static observer, and 24% for a moving observer. In far space, this weight reduced from around 5% to less than 1% at 100 m.

It has often been assumed that binocular cues only make a contribution to depth perception in near space. For example Gregory suggested that they are ineffective beyond 6 m (Gregory, 1973). In contrast, we calculated that binocular disparity should allow relative depth judgements up to a distance of more than 500 m. We also showed that, even in typical environments with multiple other cues available, binocular vision should improve depth judgements at distances of beyond 30 m.

Our estimate of the maximum useful distance for binocular cues is smaller than the 1 km calculated by Palmisano et al. (2010). This difference reflects the assumed sensitivity to binocular disparity; while we used a value of 20 arc sec (Cutting & Vishton, 1995), Palmisano et al. assumed a greater sensitivity of 10 arc sec. In both cases, binocular vision is expected to contribute to distance judgements across a range of hundreds of metres. Consistent with this, disparity has been shown to be used in depth judgements up to separations of at least 248 m (Palmisano et al., 2010).

4.1.3. Motion

When an occlusion cue was available, the weight attributed to motion parallax reduced from 8% at 2 m to 0% at 30 m. When occlusion was not available, this weight was 37% at 2 m, 13% at 30 m, and 2% at 100 m. Motion parallax is thus assumed to make a contribution to depth estimation across a similar distance range as binocular disparity. This reflects the assumption that sensitivity to motion parallax is similar to binocular disparity (Rogers & Graham, 1982), and contributes in a similar way to the perception of depth (von Helmholtz, 1867/1924).

4.1.4. Other cues

Although accommodation and focus cues undoubtedly contribute to the perception of depth (Held et al., 2012; Mather, 1996, 1997; Swenson, 1932; Watt, Akeley, Ernst, & Banks, 2005), our calculations show that their contribution to depth sensitivity is negligible (weighted

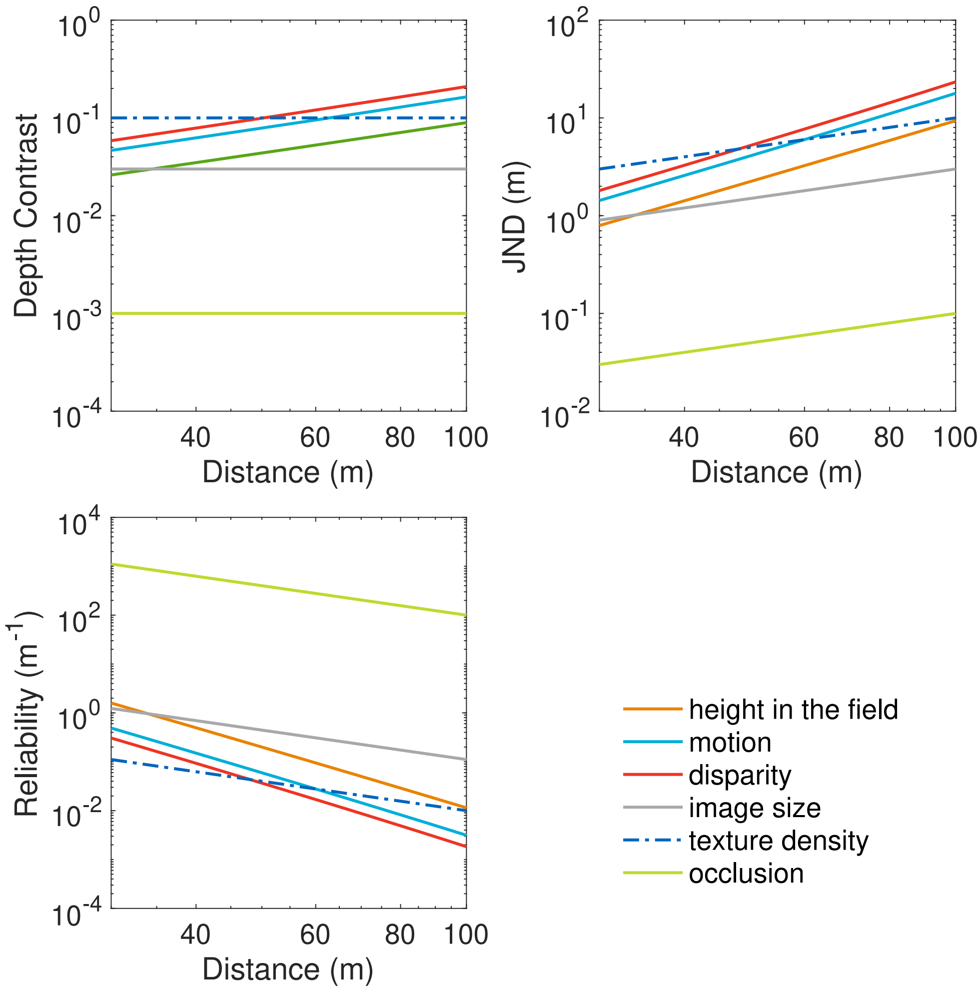

Fig. 14. (a) Depth contrasts; (b) JNDs (c) reliabilities of cues in far space, plotted in the same way as in Fig. 3.

at less than 0.2%) when other cues are available. This is consistent with the idea that they provide a complementary cue to binocular disparity, being more useful for larger differences in distance (Held et al., 2012).

Similarly, we calculated that vergence will make a negligible contribution (less than 0.2%) to sensitivity to relative depth differences. This reflects its much reduced reliability in comparison with other cues, particularly binocular disparity, which will typically accompany vergence but is much more precise. While vergence may not contribute to our sensitivity to ordinal depth judgements when other cues are available, it does play an important role in calibrating our perception of depth from binocular disparity (Bradshaw, Glennerster, & Rogers, 1996; Mon-Williams, Tresilian, & Roberts, 2000; Rogers & Bradshaw, 1995).

The other pictorial cues considered, of object and texture size and texture density, play a negligible role in relative depth judgements in near space, but a substantial role at far distances. In the absence of occlusion, the combined weight of these two cues rises from 1% in near space, through 40% at intermediate distances, to 90% at 100 m. Similar to occlusion, this reflects the fact that the reliability of these cues is uniform, while that of parallax cues falls sharply with distance.

4.2. Implementing robust and flexible cue weightings

The shear number of potential depth cues, and the variation in their availability and reliability between scenes, and across different locations within a scene, mean that weights must be assigned via a process that takes account of this complexity. This can be achieved through a process of cue combination in which weightings are assigned implicitly, rather than through the explicit calculation and continual updating of weights. A simple neural mechanism for this process has been proposed (Deneve, Latham, & Pouget, 2001). For unbiased Gaussian estimators, cue-weighting is equivalent to the multiplication of the likelihood functions for each cue. This calculation can be performed by the summation of populations of neural responses, each of which encodes the likelihood function for an individual cue (Ma, Beck, Latham, & Pouget, 2006). Two key advantages of this approach in the current context are the straightforward way in which it will handle missing or uninformative cues, and variations in the reliability of cues. When a cue is absent or uninformative, its likelihood function will be flat and it will have no influence on the combined-cue estimate. It is therefore not necessary to explicitly exclude uninformative cues, or allocate them zero weight. Similarly, since this approach to cue combination involves the combination of the appropriate population responses for every estimate made, it naturally accommodates variations in the optimal weights. It is thus able to accommodate differences in weight across conditions (Hillis et al., 2004) or across different locations in the scene (Keefe et al., 2011), without an explicit calculation and updating of these weights.

4.3. Empirical testing

The strength of many empirical tests of cue-combination has relied on the experimenter's ingenuity and precision in isolating and manipulating individual cues (Ernst & Banks, 2002; Hillis et al., 2004;

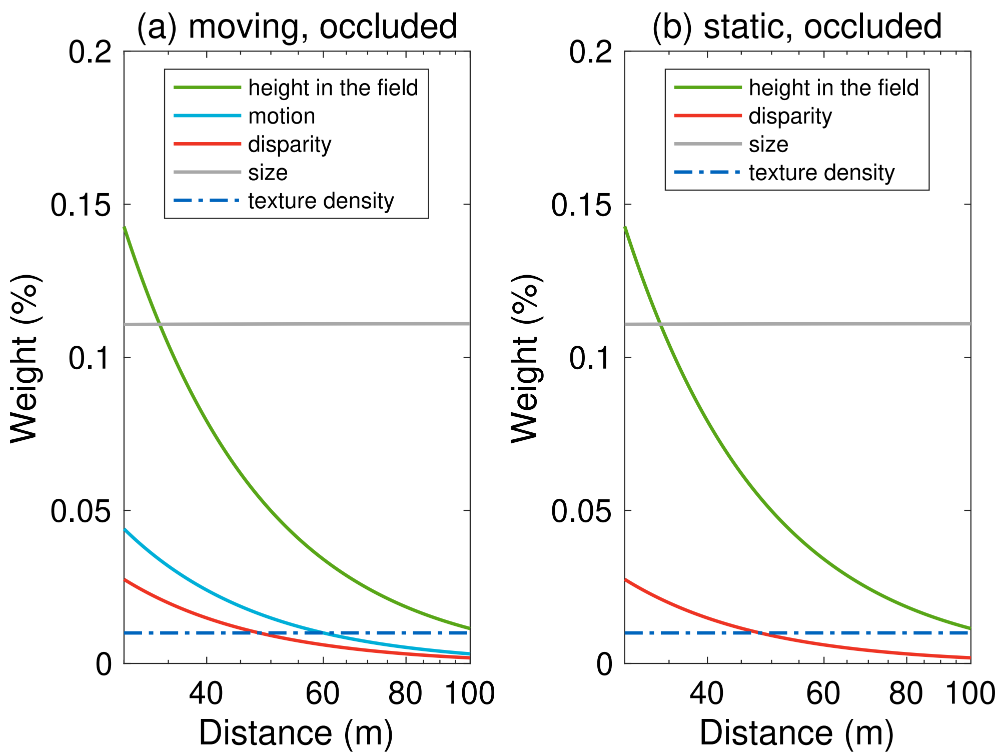

Fig. 15. Cue weights for the far distance range for (a) a moving observer and (b) a static observer, when occlusion is available. The weight for occlusion is not plotted, since it dominates other cues within this distance range.

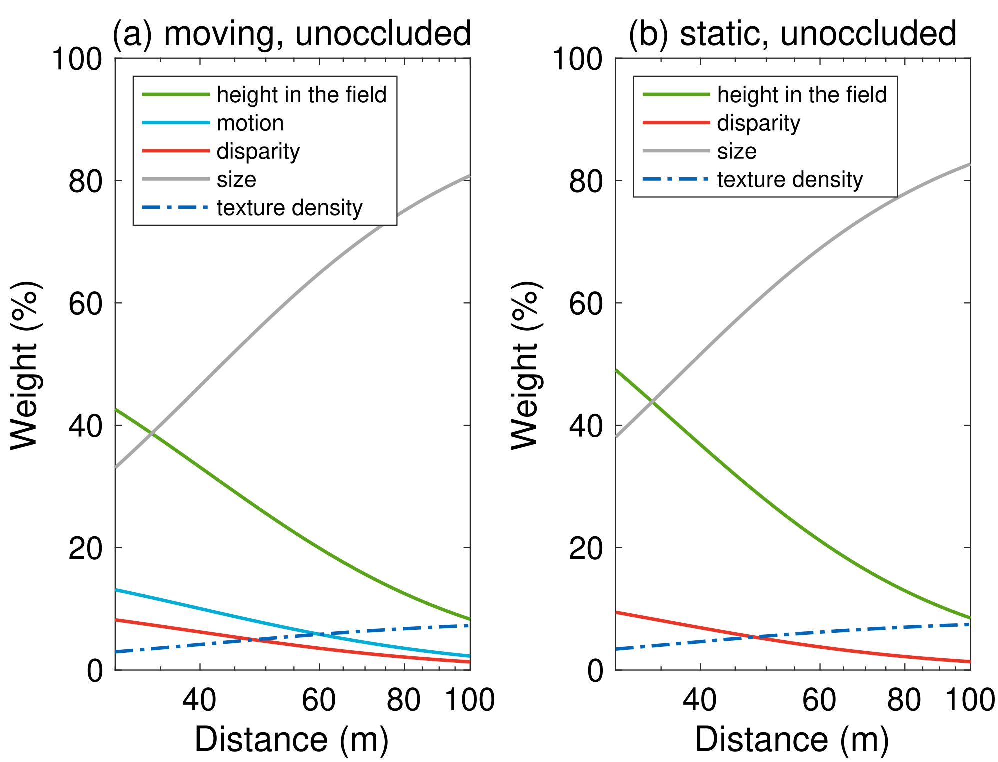

Fig. 16. Cue weights for the far distance range for (a) a moving observer and (b) a static observer, when occlusion is not available.

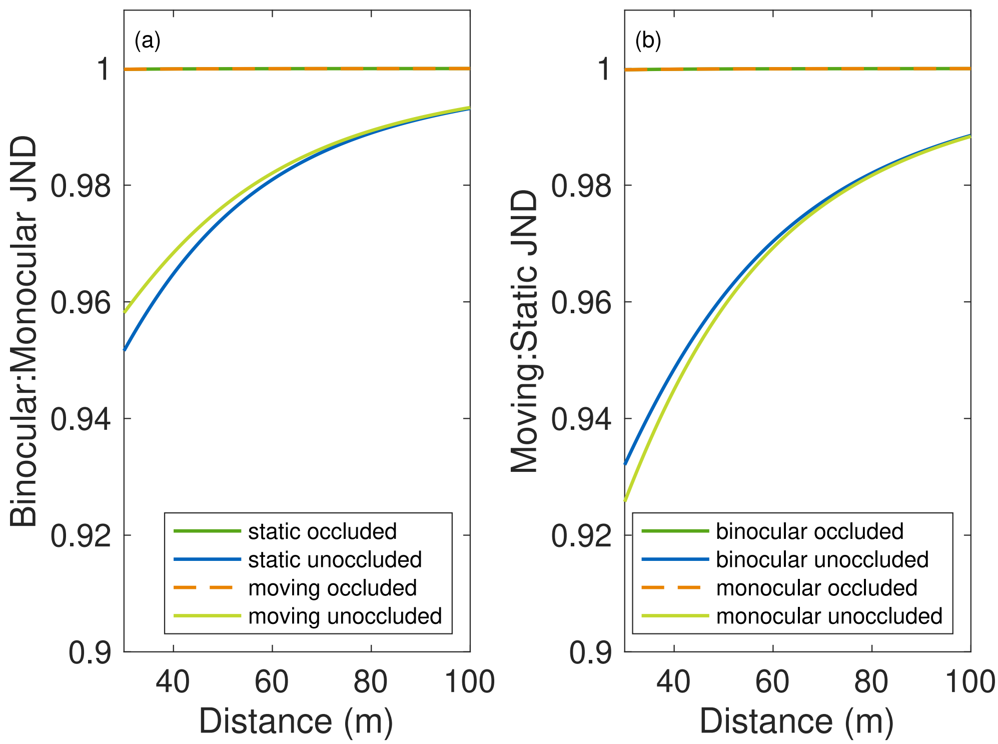

Fig. 17. The reduction in JND provided by (a) binocular cues; and (b) motion parallax in far space. JNDs were calculated as function of distance taking all cues into account, and when excluding binocular or motion cues. The improvement in JND is plotted as the ratio of the moving and static JNDs.

Keefe et al., 2011). As outlined in the introduction, while this approach provides a powerful test of the mechanisms of cue-combination, it does not allow us to quantify the weights that are likely to be assigned to individual cues in everyday vision. Our primary goal was to demonstrate that existing estimates of cue reliabilities can be used to quantify the expected weightings of these cues in complex environments containing many such cues. However, it should be noted that these reliability estimates are typically taken in very reduced cue contexts, in order to isolate individual cues, and that this approach may affect the estimated values. This may occur because, in many cases, it is not possible to remove cues, and instead, those that are not of current interest are set to fixed values. For example, in estimating the reliability of binocular vergence and disparity as cues to distance, the experimenter might present stimuli at eye height and with a constant angular size, to remove any systematic relationship between these two cues and distance. This, however, disrupts the typical statistical relationships found in most environments, which may influence observers' judgements and thus the estimates of reliability.

An additional important empirical consideration is how weightings may vary across the population. In addition to the effects of height and interpupillary distance considered here, there will be other variability in the reliabilities of individual cues across the population. Measurements of this variability, and in particular the correlations between different tasks, can provide important insights regarding the nature of perceptual mechanisms (Mollon, Bosten, Peterzell, & Webster, 2017; Peterzell & Kennedy, 2016), including those involved in the perception of depth (Harris, Chopin, Zeiner, & Hibbard, 2012; Hibbard, Bradshaw, Langley, & Rogers, 2002; Nefs, O'Hare, & Harris, 2010; Ranson, Scarfe, van Dam, & Hibbard, 2025; Wilmer, 2008). In the case of cue-combination, it is the relative reliabilities between different cues, and their independence, that will determine their weightings. This would require detailed knowledge of the full covariance matrix across the population, and within individuals. While this will provide important understanding of individual differences, such large-scale data are unavailable.

4.4. Limitations

Our approach in developing this model has been to use the estimates of cue reliability presented by Cutting and Vishton (1995) and to explore the implications of these estimates for cue weightings in typical, multi-cue situations. However, we recognise that individual reliability estimates will, in practice, depend on many stimulus parameters. For example the separation between the points (2, 5, or 10 degrees) compared by McCann et al. (2018) can influence depth discrimination thresholds. Other factors, such as image eccentricity (Hillis et al., 2004), low-level stimulus properties such as contrast and surface texture, and the speed of observer motion can also affect cue reliability. Interactions between depth cues can influence reliability and the way that cues are combined. For example, cue promotion rather than simple weighted averaging has been reported for some cue combinations (Johnston, Cumming, & Landy, 1994), while binocular disparity affects how luminance contrast is interpreted as either a depth or reflectance cue (Hibbard, Goutcher, Hornsey, Hunter, & Scarfe, 2023), and binocular viewing may reduce the ambiguity inherent in the interpretation of pictorial cues (Hibbard, Hornsey, & Asher, 2022). These findings highlight the need to consider cue dependencies when modelling depth perception.

It is also to be expected that, in practice, cues will not be independent. Correlated errors between cues will alter the optimal weights, and reduce the predicted reliability of the combined cue estimate (Oruç, Maloney, & Landy, 2003). There may also be interactions between cues that alter the availability of information. For example, occlusion of one surface by another may create a sharp boundary between one

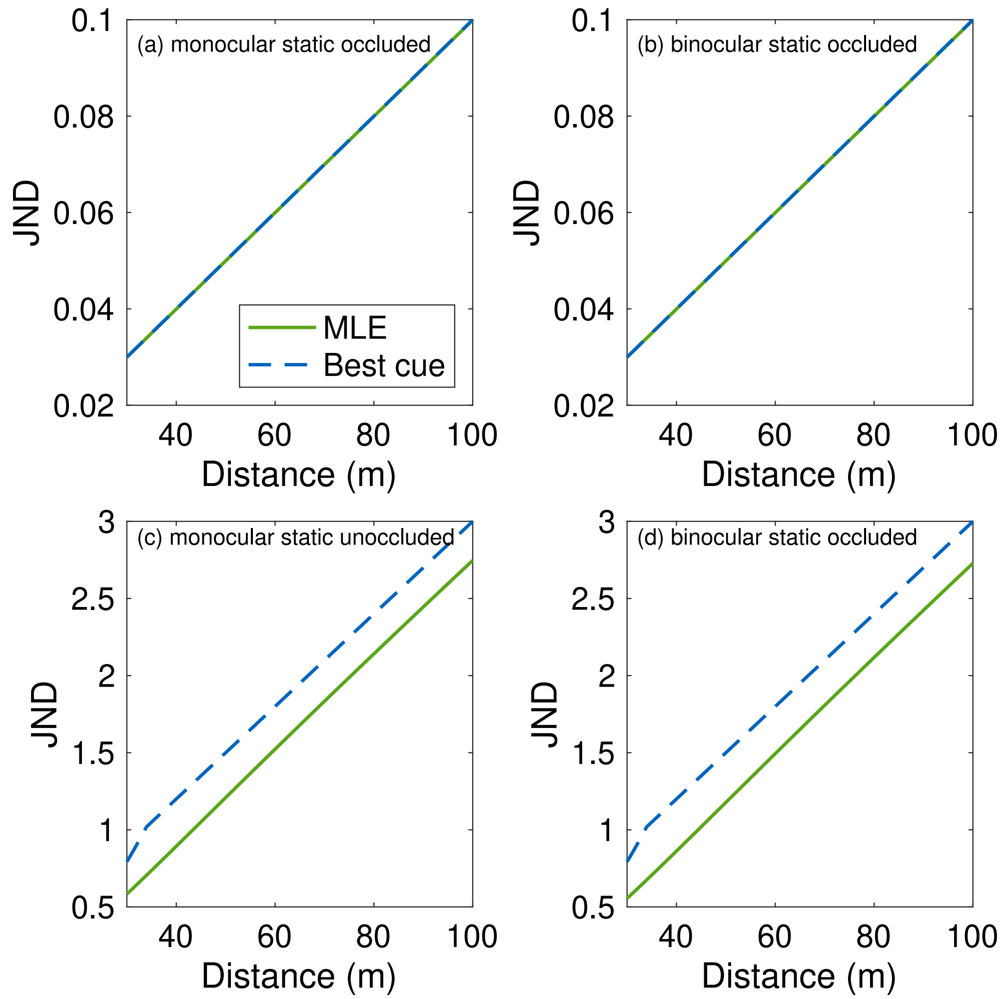

Fig. 18. JNDs in far space based on a Maximum Likelihood model (MLE) and a decision rule in which depth is estimated by using only the most reliable cue, with the smallest standard deviation (MS), for a static observer in far space. (a) Monocular viewing with occlusion (b) Binocular viewing with occlusion (c) Monocular viewing with no occlusion (d) Binocular viewing with occlusion. A clear benefit is observed for unoccluded points.

region and another, and introduce unmatched features between the two eyes (Cook & Gillam, 2004; Nakayama & Shimojo, 1990), thus violating the assumptions of the cross-correlation mechanisms underlying binocular depth estimation (Allenmark & Read, 2011; Banks, Gepshtein, & Landy, 2004). While the purpose of the current analysis is to provide an overall framework that includes this simplifying assumption, in applications where it is possible to quantify these effects they should be taken into account.

We did not include motion cues in near space, up to 2 m. Our reasoning here was that in intermediate and far space, motion parallax from a walking observer or moving objects will be an important cue, whereas in near space it will be a very common situation to be stationary and interacting with stationary objects. However, motion parallax will be generated by movement of the observer or other objects, and this could be included within extensions of the model.

4.5. Framework

Despite these limitations, our approach provides a framework in which these specific considerations can be incorporated. Rather than assuming fixed cue reliabilities, our framework allows for the integration of bespoke reliability estimates measured under conditions that reflect the specific task and viewing context (Koenderink, 1998). Future work could extend this approach by incorporating cue reliability estimates that account for variations in stimulus properties, ensuring that predictions more accurately reflect natural viewing conditions.

An alternative to the cue-isolation approach is to use cue-rich stimuli while selectively manipulating a single cue, keeping all others constant (Koenderink, 1998). This may be done, for example, by manipulating the effective gain on the relationship between depth and binocular disparity while keeping other pictorial cues constant. This approach may provide more plausible estimates of reliability that could be incorporated into the framework outlined here. Virtual reality provides a promising experimental platform for adopting this approach (Hibbard, 2023). This highlights the importance of using natural, real-world stimuli while maintaining the ability to manipulate individual cues (Koenderink, 1998; McCann et al., 2018). This approach takes advantage of naturally occurring changes in reliability based on scene parameters such as distance and viewing conditions,

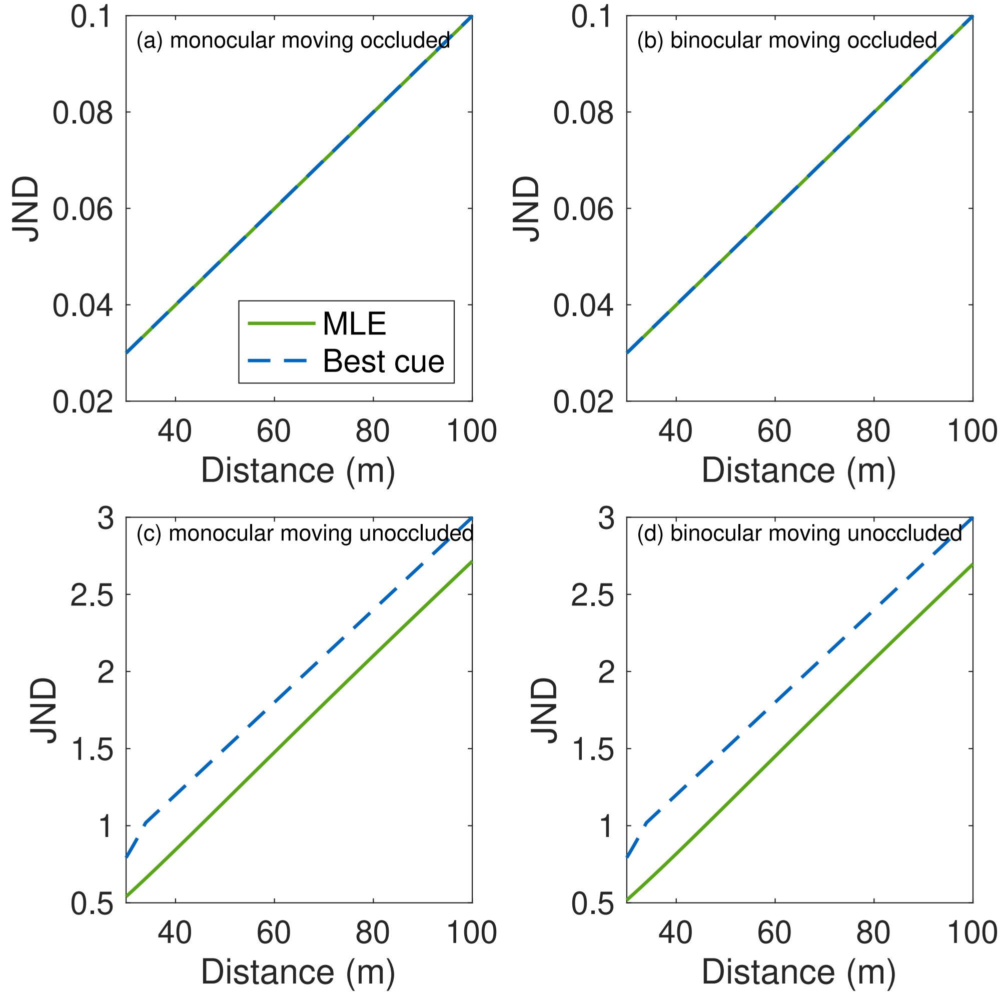

Fig. 19. JNDs in far space based on a Maximum Likelihood model (MLE) and a decision rule in which depth is estimated by using only the most reliable cue, with the smallest standard deviation (MS), for a moving observer in far space. (a) Monocular viewing with occlusion (b) Binocular viewing with occlusion (c) Monocular viewing with no occlusion (d) Binocular viewing with occlusion. A clear benefit is observed for unoccluded points.

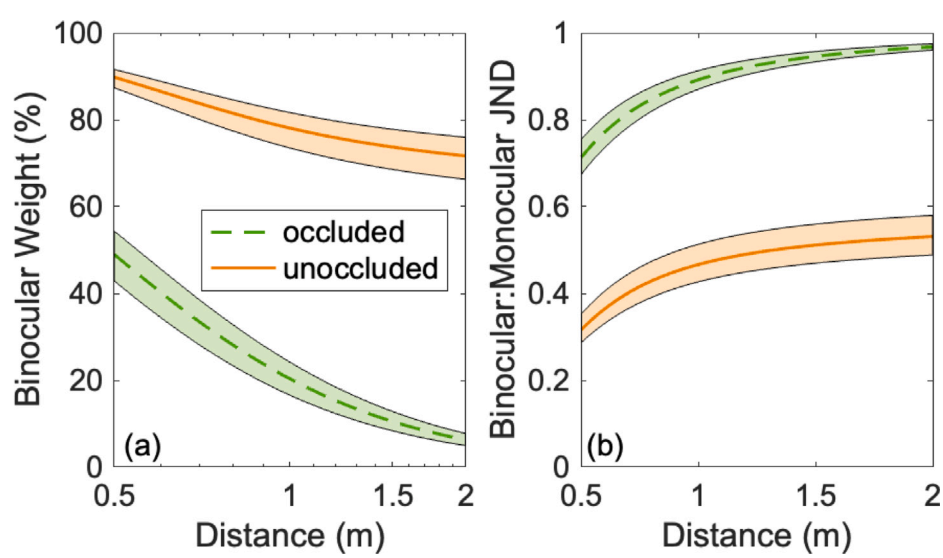

Fig. 20. Simulating the effects of changes in IPD for near space (a) The effect of distance on binocular weight across different conditions. Binocular weight (%) decreases with increasing distance, depending on if the target is occluded or unoccluded. Shaded regions show ±2 standard deviations of variation in the population IPD. (b) Binocular Just Noticeable Difference (JND) improvement as a function of distance. The ratio of binocular to monocular JND increases with distance, showing greater improvements in depth sensitivity under binocular conditions. Shaded regions show ±2 standard deviations.

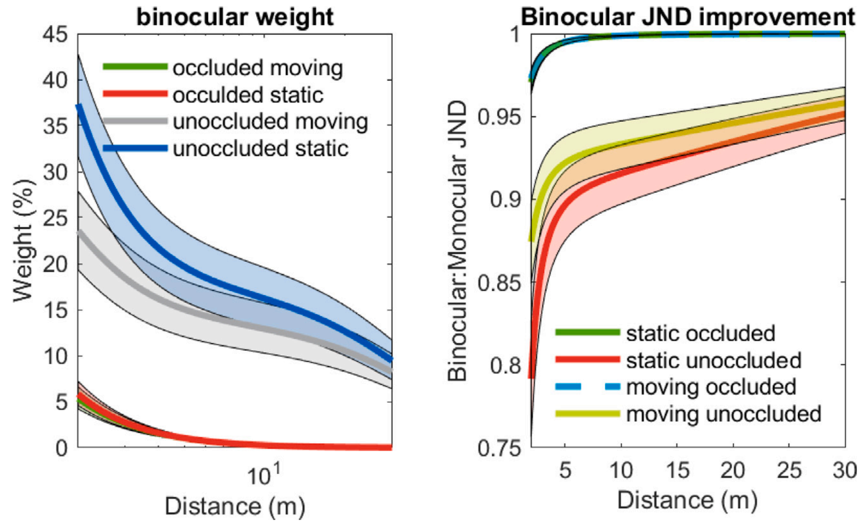

Fig. 21. Simulating the effects of changes in IPD for intermediate space (a) The effect of distance on binocular weight across different conditions. Binocular weight (%) decreases with increasing distance, depending on if the target is occluded or unoccluded and whether the observer is moving or still. Shaded regions show ±2 standard deviations of variation in the population IPD. (b) Binocular Just Noticeable Difference (JND) improvement as a function of distance. The ratio of binocular to monocular JND increases with distance, showing greater improvements in depth sensitivity under binocular conditions. Shaded regions show ±2 standard deviations.

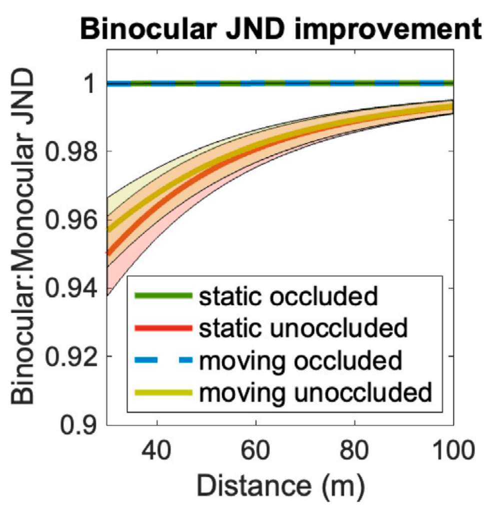

Fig. 22. Simulating the effects of changes in IPD for far space. Shaded regions show ±2 standard deviations of variation in the population IPD. In this region, binocular cues provide very little improvement in performance.

including motion speed and the availability of binocular cues, enabling testable predictions about the reliability of depth judgements.

4.6. Benefits of cue fusion

The averaging of depth cues with appropriate weights increases the precision of estimates. The magnitude of this improvement depends on the relative reliability of the individual cues. If one cue is much more reliable than the others then fusion will provide very little benefit over basing estimates on the most reliable cue alone (Rohde et al., 2016; Scarfe, 2022). This can be seen here in those cases where occlusion is available as a relative depth cue, where there is no improvement in sensitivity provided by cue fusion. In all other cases, however, cue fusion is predicted make a clear contribution to increasing the precision of relative depth estimates.

4.7. Application to other tasks

The current analysis looked only at sensitivity to depth differences, using the estimates of sensitivity provided by Cutting and Vishton (1995). The same approach can be applied to other stimulus dimensions, such as slant (Hillis et al., 2004) and object size (Ernst & Banks, 2002). In studies of cue fusion, it is typically assumed that individual estimates are unbiased, thus precluding the investigation of the biases that are known to occur (Scarfe & Hibbard, 2011; Tyler, 2020).

5. Conclusion

This study estimated the likely contribution of a range of cues to depth judgements in natural scenes. This analysis showed that the fusion of cues through weighted averaging would improve the precision of depth judgements across the whole range of distances studied, from 0.5 to 100 m. Considering the cues that are likely to be available in the natural environment, and their reliabilities, allowed us to quantify the contribution of each cue, and how this is affected by distance. These calculations indicated that the contributions of the parallax cues of binocular vision and motion parallax extend into vista space, well beyond the distance of 30 m.

CRediT authorship contribution statement

Paul B. Hibbard: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Writing – original draft. **Jordi M. Asher:** Conceptualization, Methodology, Visualization, Writing – review & editing. **Rebecca L. Hornsey:** Conceptualization, Writing – review & editing.

Acknowledgements

This research was funded by grant number RPG-2016-361 from the Leverhulme Trust, United Kingdom.

Data availability

Data will be made available on request.

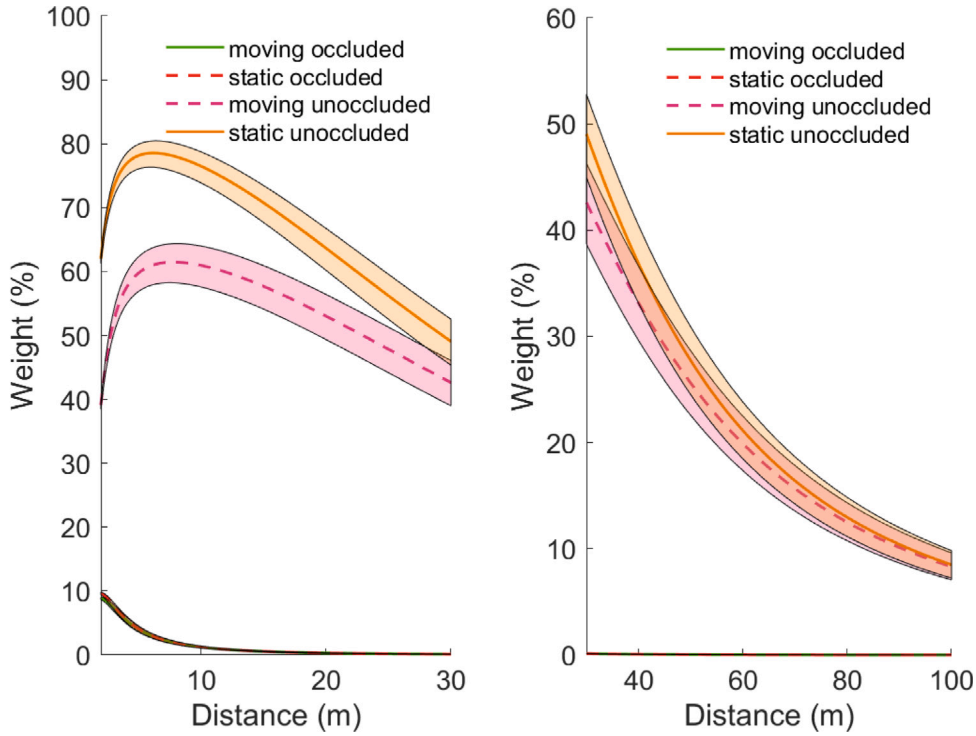

Fig. 23. Simulating the effects of observer height on the importance of the height-in-the-field cue. Weights are shown as a function of distance for intermediate (left) and far (right) space. Shaded regions show ±2 standard deviations of variation in the population height.

References

Allenmark, F., & Read, J. C. (2011). Spatial stereoresolution for depth corrugations may be set in primary visual cortex. *PLoS Computational Biology*, *7*(8), Article e1002142.

Aloimonos, J. (1988). Shape from texture. *Biological Cybernetics*, *58*(5), 345–360.

Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience*, *24*(9), 2077–2089.

Blini, E., Desoche, C., Salemme, R., Kabil, A., Hadj-Bouziane, F., & Farnè, A. (2018). Mind the depth: Visual perception of shapes is better in peripersonal space. *Psychological Science*, *29*(11), 1868–1877.

Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, *36*(9), 1255–1264.

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure–ground cue of convexity affects human depth perception. *Journal of Neuroscience*, *30*(21), 7269–7280.

Byrne, P. A., & Henriques, D. Y. (2013). When more is less: Increasing allocentric visual information can switch visual–proprioceptive combination from an optimal to sub-optimal process. *Neuropsychologia*, *51*(1), 26–37.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society*, *4*(1), 102–118.

Cook, M., & Gillam, B. (2004). Depth of monocular elements in a binocular scene: the conditions for da vinci stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 92.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In *Perception of Space and Motion* (pp. 69–117). Netherlands: Elsevier.

De Winkel, K. N., Katliar, M., Diers, D., & Bülthoff, H. H. (2018). Causal inference in the perception of verticality. *Scientific Reports*, *8*(1), 5483.

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, *4*(8), 826–831.

Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. In *Stereoscopic Displays and Virtual Reality Systems XI: Vol. 5291*, (pp. 36–46). International Society for Optics and Photonics.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Gibson, J. J. (1950). The perception of the visual world.

Goodale, M., & Milner, D. (2013). *Sight Unseen: An exploration of conscious and unconscious vision*. Oxford: OUP.

Graziano, M. S., & Cooke, D. F. (2006). Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, *44*(6), 845–859.

Green, D., & Swets, J. (1974). *Signal Detection Theory and Psychophysics* (Rev. ed.). Huntington, NY: RF Krieger.

Gregory, R. L. (1973). *Eye and Brain: The psychology of seeing*. New York City, New York: McGraw-Hill.

Harris, J. M., Chopin, A., Zeiner, K., & Hibbard, P. B. (2012). Perception of relative depth interval: systematic biases in perceived depth. *Quarterly Journal of Experimental Psychology*, *65*(1), 73–91.

Held, R. T., Cooper, E. A., & Banks, M. S. (2012). Blur and disparity are complementary cues to depth. *Current Biology*, *22*(5), 426–431.

Hibbard, P. B. (2023). Virtual reality for vision science. In *Virtual Reality in Behavioral Neuroscience: New Insights and Methods* (pp. 131–159). Springer.

Hibbard, P. B., Bradshaw, M. F., Langley, K., & Rogers, B. J. (2002). The stereoscopic anisotropy: individual differences and underlying mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 469.

Hibbard, P. B., Goutcher, R., Hornsey, R. L., Hunter, D. W., & Scarfe, P. (2023). Luminance contrast provides metric depth information. *Royal Society Open Science*, *10*(2), Article 220567.

Hibbard, P. B., Hornsey, R. L., & Asher, J. M. (2022). Binocular information improves the reliability and consistency of pictorial relief. *Vision*, *7*(1), 1.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 1–1.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, *34*(17), 2259–2275.

Jones, P. R. (2016). A tutorial on cue combination and signal detection theory: Using changes in sensitivity to evaluate how observers integrate sensory information. *Journal of Mathematical Psychology*, *73*, 117–139.

Kaas, A. L., & Mier, H. I. v. (2006). Haptic spatial matching in near peripersonal space. *Experimental Brain Research*, *170*(3), 403–413.

Keefe, B. D., Hibbard, P. B., & Watt, S. J. (2011). Depth-cue integration in grasp programming: no evidence for a binocular specialism. *Neuropsychologia*, *49*(5), 1246–1257.

Koenderink, J. J. (1998). Pictorial relief. *Philosophical Transactions of the Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences*, *356*(1740), 1071–1086.

Làdavas, E. (2002). Functional and dynamic properties of visual peripersonal space. *Trends in Cognitive Sciences*, *6*(1), 17–22.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*(3), 389–412.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438.

Makin, T. R., Holmes, N. P., Brozzoli, C., & Farnè, A. (2012). Keeping the world at hand: rapid visuomotor processing for hand–object interactions. *Experimental Brain Research*, *219*(4), 421–428.

Marotta, J., Behrmann, M., & Goodale, M. (1997). The removal of binocular cues disrupts the calibration of grasping in patients with visual form agnosia. *Experimental Brain Research*, *116*(1), 113–121.

Mather, G. (1996). Image blur as a pictorial depth cue. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *263*(1367), 169–172.

Mather, G. (1997). The use of image blur as a depth cue. *Perception*, *26*(9), 1147–1158.

McCann, B. C., Hayhoe, M. M., & Geisler, W. S. (2018). Contributions of monocular and binocular cues to distance discrimination in natural scenes. *Journal of Vision*, *18*(4), 12–12.

Melmoth, D. R., Finlay, A. L., Morgan, M. J., & Grant, S. (2009). Grasping deficits and adaptations in adults with stereo vision losses. *Investigative Ophthalmology & Visual Science*, *50*(8), 3711–3720.

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

Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research*, *141*, 4–15.

Mon-Williams, M., Tresilian, J. R., & Roberts, A. (2000). Vergence provides veridical depth perception from horizontal retinal image disparities. *Experimental Brain Research*, *133*(3), 407–413.

Moody, A., et al. (2013). Adult anthropometric measures, overweight and obesity. *Health Survey for England*, *1*, 1–39.

Nagata, S. (1991). How to reinforce perception of depth in single two-dimensional pictures. In *Pictorial Communication in Real and Virtual Environments* (pp. 553–573). Boca Raton, Florida: CRC Press.

Nakayama, K., & Shimojo, S. (1990). Da vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research*, *30*(11), 1811–1825.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology*, *18*(9), 689–693.

Nefs, H. T., O'Hare, L., & Harris, J. M. (2010). Two independent mechanisms for motion-in-depth perception: Evidence from individual differences. *Frontiers in Psychology*, *1*, 155.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451–2468.

Palmisano, S., Gillam, B., Govan, D. G., Allison, R. S., & Harris, J. M. (2010). Stereoscopic perception of real depths at large distances. *Journal of Vision*, *10*(6), 19–19.

Peterzell, D. H., & Kennedy, J. F. (2016). Discovering sensory processes using individual differences: A review and factor analytic manifesto. *Electronic Imaging*, *28*, 1–11.

Previc, F. H. (1998). The neuropsychology of 3-D space.. *Psychological Bulletin*, *124*(2), 123.

Ranson, R. E., Scarfe, P., van Dam, L. C., & Hibbard, P. B. (2025). Depth constancy and the absolute vergence anomaly. *Vision Research*, *226*, Article 108501.

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, *24*(2), 155–179.

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, *22*(2), 261–270.

Rohde, M., van Dam, L. C., & Ernst, M. O. (2016). Statistically optimal multisensory cue integration: a practical tutorial. *Multisensory Research*, *29*(4–5), 279–317.

Sambo, C. F., Forster, B., Williams, S., & Iannetti, G. (2012). To blink or not to blink: fine cognitive tuning of the defensive peripersonal space. *Journal of Neuroscience*, *32*(37), 12921–12927.

Scarfe, P. (2022). Experimentally disambiguating models of sensory cue integration. *Journal of Vision*, *22*(1), 5–5.

Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision*, *11*(7), 12–12.

Serwe, S., Drewing, K., & Trommershäuser, J. (2009). Combination of noisy directional visual and proprioceptive information. *Journal of Vision*, *9*(5), 28–28.

Sousa, R., Brenner, E., & Smeets, J. B. (2011). Judging an unfamiliar object's distance from its retinal image size. *Journal of Vision*, *11*(9), 10–10.

Sousa, R., Smeets, J. B., & Brenner, E. (2012). Does size matter? *Perception*, *41*(12), 1532–1534.

Swenson, H. A. (1932). The relative influence of accommodation and convergence in the judgment of distance. *The Journal of General Psychology*, *7*(2), 360–380.

Taya, S. (2023). Do observers use their own interpupillary distance in disparity scaling? *Optical Review*, 1–9.

Todd, J. T. (2020). On the ambient optic array: James gibson's insights about the phenomenon of chiaroscuro. *i-Perception*, *11*(5), Article 2041669520952097.

Trommershäuser, J., Körding, K., & Landy, M. S. (2011). *Sensory Cue Integration*. Oxford: Oxford University Press.

Tyler, C. W. (2020). An accelerated cue combination principle accounts for multi-cue depth perception. *Journal of Perceptual Imaging*, *3*(1), 10501-1.

von Helmholtz, H. (1867/1924). *Treatise on Physiological Optics, 3 vols*. Optical Society of America.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, *5*(10), 7–7.

Watt, S. J., & Bradshaw, M. F. (2000). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia*, *38*(11), 1473–1481.

Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision*, *21*(6), 561.