Citation: Linton, P. (2020). Do we see scale?. (Unpublished Doctoral thesis, City, University of London)

DO WE SEE SCALE?

by

PAUL LINTON

A dissertation submitted for the Degree of Doctor of Philosophy

City, University of London

Division of Optometry and Visual Sciences

Visual Neuroscience and Psychophysics Research Group

July 2020

Abstract

Since Ibn al-Haytham (c.1021), one of the central questions of vision science has been how the visual system extracts absolute distance information from the environment in order to scale the size and distance of objects in the scene. Since Kepler (1604) and Descartes (1637), emphasis has been placed on a number of distinct 'triangulation' cues: vergence (the angular rotation of the eyes), accommodation (the focal power of the eyes), motion parallax (the change of the scene with motion), and, more recently, vertical disparities (the differences in the vertical projection of points in the two eyes). However, there is increasing skepticism about accommodation and motion parallax, and I argue that when vertical disparities have been properly tested, they too prove ineffective. Which only leaves vergence. However, in three experiments I demonstrate that vergence is not an effective absolute distance cue. Which leads us to the following conclusion: If all of the remaining potential absolute distance cues are either cognitive in nature (familiar size) or limited in context (ground plane), we should instead be open to the idea that the visual system operates without absolute distance information, and scale is not conveyed by our visual experience.

[Please cite published / preprint work rather than this thesis]

Acknowledgements

In 2015, I was reading (1) Chen, C.C. and Tyler, C.W. (2015). 'Shading beats binocular disparity in depth from luminance gradients: Evidence against a maximum likelihood principle for cue combination', *PLoS ONE*, 10(8), and (2) Melmoth, D.R. and Grant, S. (2006). 'Advantages of binocular vision for the control of reaching and grasping', *Exp Brain Res*, 171(3), pp. 371–388 alongside one another, and was struck that Christopher Tyler and Simon Grant were in the same department. Though we have very different approaches to 3D vision, our common concerns – the validity of Bayesian cue integration, the significance of binocular vision – are evident in those papers. It has been an absolute pleasure to be supervised by Christopher and Simon, and it's been a privilege to discuss these questions with them, particularly during Christopher's stays in London.

Another massive debt has been to Chris Hull, who very kindly afforded me space in his Ocular Optics Lab to perform the three experiments in this PhD, as well as invaluable technical assistance. It also gave me the immense benefit of sharing a hallway with Michael Morgan, Joshua Solomon, Matteo Lisi, and Pete Jones, and it has been wonderful to pop my head in and discuss vision science with them. Michael, Josh, Matteo, and Pete have all been very generous with their time and advice, and I really appreciate it. Matteo, in particular, has had some success in his insistence that I embrace R, and begin on a journey of hierarchical Bayesian modelling.

Salma Ahmad was absolutely amazing as the optometrist on the third experiment. Thank you to John Barbur and Ron Douglas for help on the first experiment (testing whether the red filter used affected accommodation). And thank you to Mark Mayhew, Deanna Taylor, and Jugjeet Bansal, for fitting contact lenses during piloting of the third experiment, and Byki Huntjens and Ahalya Subramanian for their advice, as well as Priya Mehta and the CitySight clinic for their help.

The wider London vision science community (Tessa Dekker, John Greenwood), and the wider UK 3D vision community have been extremely welcoming and supportive. In particular Andrew Glennerster, Jenny Read, and Dhanraj Vishwanath, have been extraordinarily generous with their time. It is notable that the four of us came to 3D vision from other disciplines (Andrew: medicine, Jenny: astrophysics, Dhanraj: architecture, and me: law / philosophy). And I greatly appreciated the support of Fulvio Domini, Thomas Papathomas, and Peter Tse in the United States.

I also had a hugely enjoyable four months as part of the DeepFocus team at Facebook Reality Labs in September to December 2018, working primarily with Marina Zannoli, Lei Xiao, Salah Nouri, and Douglas Lanman, with Gusti Lulu Fatima as an amazing research assistant. It was a fantastic experience, in large part due to the people there, and I must also thank Max Di Luca, Cesare Parise, Kevin MacKenzie, and James Hillis for being generous with their time and advice.

Elements of this work have previously been published as Linton (2017) and Linton (2020).

Paul Linton

July 2020

Contents

List of Figures

Outline

The argument of this thesis is that the visual system does not appear to have access to absolute distance information, and we should therefore be open to the idea that vision does not directly provide us with absolute scale. This argument rests on the following propositions:

1. The effectiveness of motion parallax is widely questioned in the literature [Ch.2]

2. Evidence supporting vertical disparities relies on misrepresenting them [Ch.2]

3. Evidence supporting vergence relies on introducing confounding cues [Ch.2]

4. Vergence is not an effective absolute distance cue once these confounding cues have been controlled for [Ch.4]

5. This ineffectiveness cannot be attributed to vergence / accommodation conflict [Ch.5]

6. Nor is vergence an effective cue to size constancy [Ch.8]

7. Interim Conclusion: This suggests that the visual system primarily relies on 'cognitive' cues to absolute distance such as familiar size [Ch.9]

8. The problem with this conclusion is the 'paradox of visual scale': Triangulation cues are ineffective, and yet we know from Telestereoscopic viewing that triangulation cues dominate pictorial cues such as familiar size [Ch.9]

9. The solution is to suggest that binocular scale is gauged by comparing the perceived 3D shape from binocular disparity with the perceived shape either (a) from monocular cues, or (b) from natural scene statistics [Ch.9].

1. The Problem of Visual Scale

Visual scale and size constancy are two related problems in visual space perception that must be resolved if we are to effectively translate the proximal stimulus (the retinal image) back to the distal stimulus (the perceived three-dimensional object). Both of these problems are caused by the fact that the retinal image gets smaller with distance, and can be illustrated with the following diagrams:

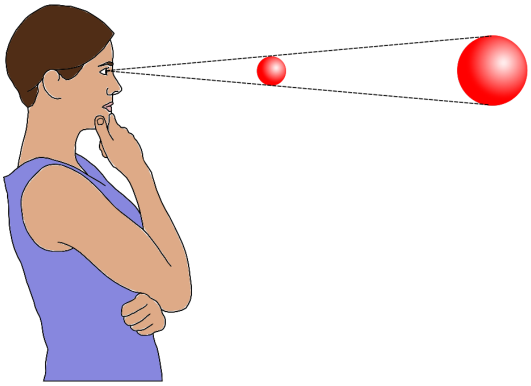

Fig.1. The Problem of Visual Scale

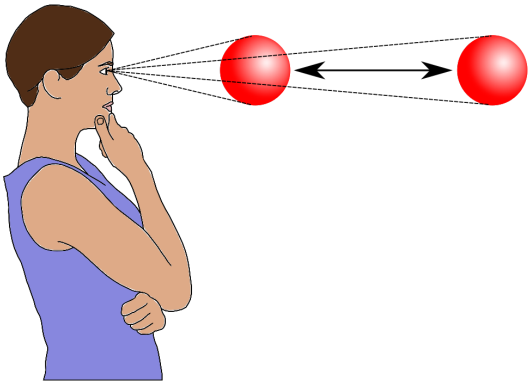

Fig.2. The Problem of Size Constancy

The problem of visual scale is how we are able to tell whether what we are looking at is a small object up close or a large object far away, given that they can both cast the same size retinal image. The problem of size constancy is that as an object moves forwards and backwards in space, its retinal image will vary drastically in size. And yet in our everyday perception, objects don't appear to vary dramatically in size as they move forwards and backwards in space. Solving both of these problems, visual scale and size constancy, relies on finding a distance mechanism that tells us how far the object is, in order to scale the retinal image depending on the object's distance.

Visual Scale has two components: Size and Distance. Specifically, as we have just noted, it is concerned with differentiating a small object up close from a large object far away, even though both may have the same visual angle. Ptolemy (c.160 AD) first articulated visual scale in these terms, arguing that size wasn't just a function of visual angle (vs. Euclid, c.300 BC), but visual angle appropriately scaled by distance information (see Hatfield, 2002), although Boring (1942) suggests some element of a realisation of this was also made by Euclid (c.300 BC):

> "[Euclid] also modified his strict geometry by noting that equal magnitudes at unequal distances do not appear in sizes inversely proportional to their distances. Thus, it appears that in moments of phenomenological insight, the distinction between apparent size and visual angle goes back to the beginning." (p.290)

However, the strict geometric relationship between perceived size and distance, the 'size-distance invariance hypothesis', posited by Ptolemy (that size is the retinal image scaled by the distance), is not without exceptions. As far back as the 1950s, Kilpatrick & Ittelson (1953) concluded that in light of the number of reported exceptions to the hypothesis, the 'size-distance invariance hypothesis' "loses its status as a powerful explanatory concept and becomes rather a description of results obtained under conditions which have yet completely to be specified."

As an illustration, one notable example of the violation of 'size-distance invariance hypothesis' is Foley et al. (2004). They asked participants in an open air field in full cue conditions (but sat down and in a chin rest, so no motion parallax) to verbally report a judgement about (a) the observer's absolute distances to 14 stakes in the ground and (b) the distances between the stakes (91 combinations, leading to 105 reports in total). Stakes varied in height from 40 to 160 cm. A reference stake 1m long was placed beside the observer. There were 4 conditions, with the stakes either (a) placed at distances 5-23m or (b) 18-36m, and the observer either viewing (a) monocularly or (b) binocularly (participants only participated in one of the four distance-range / viewing-condition combinations). Interestingly, participants were far more accurate at estimating the egocentric distance to the stakes (which Foley et al., 2004 take as the measure of 'perceived distance') than the distances between stakes (which Foley et al., 2004 take as the measure of 'perceived size'). This is illustrated by Figure 3, where predicted size (angular size scaled by perceived egocentric distance, on the x-axis) underestimates the actual size reported by participants (on the y-axis).

[Removed due to copyright]

Figure 3. The relationship between reported vs. predicted size (extent) in Foley et al. (2004) for two distance ranges (5-23m and 18-36m, indicated by 5 and 18 respectively), and two viewing conditions (monocular and binocular). The white squares indicate reported distance judgements, and illustrate a close to veridical relationship ($y = x$). By contrast, the reported size (extents), as indicated by the blue diamonds, are over-estimated ($y > x$), which Foley et al. (2004) take to be a violation of the size-distance invariance hypothesis.

Foley et al. (2004) therefore conclude that "although the perception of location and the perception of extent are related, they not related by Euclidean geometry, nor by any metric geometry." Hence, we will describe visual scale in approximate relationship between perceived size and perceived distance, for instance as the ability to effectively differentiate a small object up close vs a large object far away, rather than suggesting a strict 'size-distance invariance'.

Given we have defined visual scale in at least approximate terms, the next question is where does the distance information that the visual system needs to scale the retinal image come from? For Ptolemy, who was the first to articulate visual scale as the scaling of the visual angle, the answer was easy. As an 'extramission' theorist he could rely on rays emitted from the eyes, then returning to the eyes, using the length of these rays to estimate the distance of objects. In practice some animals (bats, dolphins, whales, and even some birds and rodents) have evolved what is effectively an 'extramission' theory of hearing to address this very concern. However, in the absence of echolocation in humans, 'intramission' theorists such as Kepler (1604) and Descartes (1637) posited a number of visual cues that might plausibly have replaced Ptolemy's 'extramission' theory of vision, most notably vergence and accommodation. The argument of this thesis is that given vergence and accommodation's demonstrated ineffectiveness in my experiments, we should conclude that the human visual system does not extract absolute distance information from the environment, and we should embrace a theory of vision without scale.

Responses to finding a source of absolute distance information consistent with 'intramission' have generally fallen into two camps: (1) pictorial, and (2) triangulation.

1. Pictorial Cues: In the middle ages the eye took on the kind of fascination that is currently reserved for the brain. Although Aristotle (c.335-323 BC) was an early advocate of the 'intramission' theory, 'intramission' was only truly established by Ibn al-Haytham (c.1021) (often Latinised as 'Alhazen'). al-Haytham (c.1021) recognised the challenge of attributing distance to a visual system based on projection, and proposed two solutions. The first was the ground-plane, and the second was familiar size [*Optics*, 3.67-3.93]. On the ground-plane, al-Haytham (c.1021) suggests that "the magnitude of the distances of visible objects is not correctly perceived by the sense of sight unless distances of visible objects are spanned by a continuous, ordered range of bodies", and gives the ground that connects two walls as an example. On familiar size, al-Haytham (c.1021) suggests that where "a continuous, ordered range of bodies" is not present, the visual system will compare the object's distance to "the distance of similar visible objects that have been perceived before by sight". Both the ground plane and familiar size are commonly thought of as pictorial cues that are present in the static monocular image. However, Descartes (1637) suggested that familiar size was merely cognitive in nature: "we may already have from another source an image of an object's size … and this may enable us to *imagine its distance, if not actually see it.*" (Emphasis added). Another pictorial cue that we will consider is defocus blur (Vishwanath & Blaser, 2010; Held et al., 2010).

2. Triangulation Cues: Instead, both Kepler (1604) and Descartes (1637) emphasised the three triangulation cues that continue to be given prominence in the literature today: vergence, accommodation, and motion parallax. Strictly speaking, Kepler (1604) didn't understand that the eye had a lens that changed in strength and so conceived of the eye as in principle akin to an aperture (like a nautilus eye), and suggested that the degree of defocus blur of a point through this aperture would triangulate its exact distance. Descartes (1637) corrected this misapprehension, and explained that the only way to extract absolute distance from light ray divergence was accommodation (putting to one side the question of whether the sign of defocus can be extracted by the visual system; i.e. whether the only information provided by the light ray vergence is simply the width of the blur circle cast on the retina). We can also add vertical disparities to the list of potential absolute distance cues following theoretical work by Longuet-Higgins (1981), Mayhew & Longuet-Higgins (1982), and Gillam et al. (1988).

The purpose of the thesis is to make two arguments, a broad argument and a narrow argument. The broad argument is that all of our absolute distance cues appear to be either (1) largely ineffective (vergence, accommodation, vertical disparities, motion parallax), (2) constrained to very limited circumstances (ground plane, visual field blur), or (3) purely cognitive in nature (familiar size). In that case, we have to question whether our visual experience really conveys a sense of absolute size and distance.

The narrow argument is what I call the 'paradox of visual scale', which is this:

1. We have separated cues to visual scale into two categories: (1) triangulation cues (vergence, accommodation, vertical disparities, motion parallax), and (2) pictorial cues (familiar size, ground plane, visual field blur).

2. I argue that all of the triangulation cues are largely ineffective.

3. But at the same time we know (from Telestereoscopic viewing by Helmholtz, 1858) that triangulation cues dominate pictorial cues so far as visual scale is concerned. This is the fact that if we use mirrors to increase the interpupillary distance (the distance between the eyes), and the observer looks at the world, "it seems as if the observer were looking not at the natural landscape itself, but at a very exquisite and exact model of it, reduced in scale" (Helmholtz, 1858, see Figs.41 and 42 below).

4. The solution to this paradox is to invert the classic relationship between distance and 3D shape. Rather than arguing that we need to extract distance to preserve 3D shape (shape constancy), I argue that we extract absolute scale from the way in which 3D shape is distorted with distance (failure of shape constancy).

2. Absolute Distance Cues

In the remainder of this introduction, I briefly survey each of these absolute distance cues in turn in order to substantiate the general account.

Although both the broad and narrow arguments are quite stark conclusions to have to come to, they actually cohere with much of the experimental literature on these individual cues. The one key exception is vergence, explaining why this will be the focus of my three experimental studies, two on vergence as a distance cue, and one on vergence as a cue to scale.

1. Familiar Size: Let us start with the contention that familiar size is a merely cognitive cue. Note that by *cognitive* I do not mean to imply that subjects have to consciously infer distance on the basis of familiar size. Instead, the inferences of 'visual cognition' can be *automatic* (we don't have to do anything), *unconscious* (we are unaware of them), and *involuntary* (we cannot overrule them), and still be *post-perceptual.* An example is reading: we do not have to *consciously* attribute meaning to words on a page, and yet it is hard to maintain that someone who understands a language and someone who does not have a different *visual* experience of the page itself. And as we have already discussed, Descartes (1637) clearly thought of familiar size along these lines ("this may enable us to *imagine its distance, if not actually see it*", emphasis added).

I agree with the current consensus which appears to favour treating familiar size merely as a cognitive cue to distance. Specifically, it is not a source of absolute distance information that the visual system uses to scale the scene. In a series of papers over 30 years, Walter Gogel (Gogel, 1969; Gogel, 1976; Gogel & Da Silva, 1987; Gogel, 1998) and John Predebon (Predebon, 1979; Predebon, 1987; Predebon, 1990; Predebon, 1992a; Predebon, 1992b; Predebon, 1993; Predebon, 1994; Predebon & Woolley, 1994) questioned whether familiar size really affects our visual perception of scale, and found (in the words of Predebon, 1992b) that "the influence of familiar size on estimates of size mainly reflects the intrusion of nonperceptual processes in spatial responses." Citing this literature Vishwanath (2014) concludes:

> "There are no studies that have conclusively demonstrated that familiar size is an independent quantitative perceptual cue to distance. The most recent consensus is that, on its own, familiar size only affects the cognitive inference of distance (Gogel & Da Silva, 1987; Predebon, 1993)."

For instance, Gogel (1976) found that familiar objects do not have the perceived motion parallax that one would predict from subject head motion had familiar size determined their distance, whilst Predebon (1992b) found that the influence of familiar size could be vastly reduced simply by asking subjects for the apparent rather than physical size of the stimulus, using the following instructions to his participants:

> "By actual or real size (distance) I mean the size (distance) you would obtain if you were to measure the (distance of the) object with a measuring tape or ruler. … By apparent size (distance) I mean the size (distance) which the object looks to you on your first visual impression without going into the reason for that impression."

It is also worth noting just how easily familiarity cues to visual scale can be negated; for instance in an Ames Room, or in the context of tilt-shift miniaturisation from visual field blur (discussed below). This is not to suggest that familiar size is a weak cue. Indeed, I would argue that familiarity, in a broad sense, is responsible for our impressions of pictorial scale, and monocular perceptions of scale. But there is still an important point here. Take a picture of an Ames room. What provides absolute size and distance in this scenario? The two options appear to be (a) familiar size, and (b) height in the ground plane. But height in the ground plane appears to function merely as a relative depth cue in this scenario. It doesn't tell us that the further person is the physically correct size for the room (which they are, according to the ground plane, since they are at ground level, with the nearest person raised in height using a false floor). So familiar size is not being extinguished by another absolute size and distance cue in the Ames illusion. Instead, the familiar size of the people is determining the room's scale. Although we might think that familiarity with the room is providing its scale, that would be a mistake. If both the people were small we would think it was a very large room. If both the people were big, a very small one. So it's the fact of the size differential between the two people that makes us average between them, and think that this is a standard sized room ≈ 2.4m. Instead, the point is that familiar size is providing scale to the scene, i.e. functioning as an absolute distance cue, and then being completely negated as a relative depth cue (in terms of understanding the relative geometry of the people in the scene), and that seems jarring unless you think of its contribution to visual perception as merely cognitive.

I will proceed on the basis that the literature is correct, and I can take as a point of departure that familiar size enables merely cognitive inferences about the size and distance of objects. But in order to both (a) to make clear what this means, and also (b) to outline the kind of experiment that would really drive the point home, let me consider the following hypothetical example. Participants monocularly view a luminous rectangle or a luminous circle. We then transform that luminous object into a familiar object (a playing card or a coin) by means of a 'Pepper's ghost' illusion that places them at the same optical distance. Replacing one (unfamiliar) object with another (familiar) one clearly gives us new distance information. The question is how this new distance information is processed? Does the new (familiar) object move forwards or backwards in space as it is perceptually relocated along the z-axis by the visual system? Or does it remain at the same position along the z-axis as the old (unfamiliar) object? (Only now the observer has a better idea of what this distance corresponds to). If it is merely the latter, then familiar size is not functioning as a visual cue; it is not being used to locate the object in the 3D visual scene.

2. Ground Plane: The ground-plane is articulated as an important source of absolute distance information by Gibson (1947); Gibson (1950); Ooi, Wu, & He (2001); Li & Durgin (2012); Tyler (2018). Blind walking to an object viewed on the ground plane has been found to be accurate for distances between 2.7m and 6.9m by Ooi & He (2015). Monocular participants were asked to blind walk to the distance of an object (styrofoam ball of constant angular size) placed on the ground in a corridor, and also to the distance of an object suspended 67cm above the ground plane (and indicate its height). Ooi & He (2015) found that distance judgements were almost perfect for objects on the ground plane, but distances were compressed towards the mean for the object suspended above the ground plane. Although it is worth noting that the distance compression for the suspended target (compare the compression in the x-axis of the black monocular circles with the + target or, where this is occluded, the white binocular circles) is not dramatic. The nearest targets are roughly veridical, and the furthest target (6.9m) is judged to be around 6.3m.

[Removed due to copyright]

Figure 4. Absolute distance judgements from Ooi & He (2015) (Top) and Foley et al. (2004) (Bottom). Top Left: Distance judgements in Ooi & He (2015) for an object on the ground plane. Top Right: Distance and height judgements in Ooi & He (2015) for a suspended object. Bottom Left: Distance judgements for a binocular observer in Foley et al. (2004). Right: Distance judgements for a monocular observer in Foley et al. (2004).

Gibson (1950), Sedgwick (1983), and Ooi & He (2015) argue for a 'ground theory of spatial perception', with Gibson (1950) going as far as to suggest that "*there is literally no such thing as a perception of space without the perception of a continuous background surface*." But the problem with this literature is either (a) it is tested in full-cue conditions, in which case there is no control for familiarity cues providing the absolute scale of the scene (and the ground plane just acts as a relative depth cue indicating where in the scene the target is) (this is the dilemma that Gibson / Sedgwick / Loomis / Ooi face), or (b) the ground plane is tested in controlled conditions, in which case it biases our judgements (nearer vs closer), rather than acting as an absolute distance cue (Li & Durgin, 2012). Gibson (1950) presents a refined, abstracted, idea of the ground plane:

[Removed due to copyright]

Figure 5. Example of the ground plane from Gibson (1950)

But it's unclear whether this abstract ideal is actually being tested by the corridor in Ooi & He (2015), or even the outdoor field in Foley et al. (2004), where there's so much contextual (familiarity) information by which to scale the scene. There's also a deeper question about familiarity. Gibson (1950) talks about "*the perception of a continuous background surface*", but what identifies this background surface as the ground plane rather than e.g. a table, other than familiarity.

What is the alternative to Gibson (1950)'s suggestion that "*there is literally no such thing as a perception of space without the perception of a continuous background surface.*" It is that the ground-plane should be thought of as a triangulation cue, rather than a pictorial cue. Specifically, that we don't need to see a continuous background surface, only know that the object is on the ground plane, and then we can triangulate its distance using (a) eye height, and (b) angle of declination.

[Removed due to copyright]

Figure 6. Sedgwick (1983)'s proposal for distance from the ground plane using pictorial cues alone. Diagram from Rand et al. (2011)

Let's assume that eye height is known. How are we to gauge angle of declination? Sedgwick (1983) suggests reliance on the horizon. This has the benefit that we have a retinal cue to angular declination, rather than a noisy vestibular one. But, first, we rarely encounter the horizon in everyday viewing. And second, how do we know that the horizon is at the same elevation that we are, and is not sloped upwards or downwards?

It therefore appears that reliance on a noisy interplay between (a) eye muscles, (b) neck muscles, (c) postural muscles (upright or stooping?), and (d) vestibular cues, appears inevitable.

These are cues that wouldn't be present in a picture, but would have to be essential if we wanted to show that the ground plane is an effective absolute distance cue in normal viewing conditions.

There's some evidence for this. For instance, Török et al. (2017) had subjects judge the distance of an object between 5m and 25m with three head inclinations: downwards by -20°, 0°, and upwards by +20° whilst keeping the retinal image fixed. They also applied left-right or right-left galvanic vestibular stimulation to see if this would accentuate the effect.

[Removed due to copyright]

Figure 7. Experimental paradigm and results from Török et al. (2017)

They found that participants tended to under-estimate distances when the head was rotated downwards by -20°, and over-estimated distances when the head was rotated upwards by +20°. And the galvanic vestibular stimulation accentuated this effect, although it is unclear why from the perspective of triangulating the distance of the object on the ground plane. As Török et al. (2017)

explain, the GVS "is considered to enhance the vestibular activity by mimicking a natural movement of the head, which elicits a virtual sensation of roll tilt". Török et al. (2017) conclude:

> "Our results support the gravity theory of depth perception, and show that vestibular signals make an on-line contribution to the perception of effort, and thus of distance."

But there are three concerns with deriving too much from these results.

First, there are conflicting results. For instance, Clément et al. (2016) found short periods of microgravity (0G) and hypergravity (1.8G) during Zero-G parabolic flight had no effect on verbal distance reports. One possibility is that whilst Clément et al. (2016) truly did keep the retinal image the same in these three conditions (because the participant was located within, and therefore saw, the same physical object), the visible seams of the CAVE system in Török et al. (2017) may have provided non-vestibular visual cues to angular declination.

Second, Török et al. (2017) demonstrate that vestibular signals make *a* contribution to distance perception, but little beyond this. The ground plane increasingly begins to take on the status of a spatial bias, rather than a metric estimate of absolute distance – objects seem nearer or further, rather than a specific distance. This is important, because the kinds of computer generated scenes in Török et al. (2017) remove the familiarity/contextual cues which I argue are confounding cues in the blind-walking studies discussed above. A mere spatial bias is how Ooi, Wu, & He (2001) and Loomis (2001) tend to discuss the ground plane, and is often referred to as the "intrinsic bias" account (Li & Durgin, 2012b). Indeed, a series of spatial biases should suffice for distances beyond reaching space, where simple heuristics indicating the approximate distance can be corrected during motion. And, as we have already discussed, any additional accuracy in blind walking paradigms, may simply be attributable to familiarity cues that are rarely controlled for in experiments.

Morgan (1989) suggests that online correction could also suffice for reaching and grasping:

> "In normal reaching behaviour we have binocular vision not only of the object we are reaching for, but also of the hand. In these circumstances, it is not clear that the primary goal of stereopsis should be the extraction of absolute rather than relative distance information. Extraction of shape may be an additional bonus of stereopsis, rather than its primary function."

But I will argue in Experiment 1 and Experiment 2 that hidden-hand pointing is very accurate, so what this suggests is that we need accuracy in reaching space, whilst we can afford to rely on rougher approximations of further distances, explaining why my thesis focuses on near distances.

Third, Török et al. (2017)'s conclusion that "vestibular signals make an on-line contribution to the perception of effort, and thus of distance" sound less like this visual system extracting absolute distance on the basis of simple pictorial geometry (Sedgwick, 1983's proposal) and closer to the discussions of embodied perception and whether perceived effort affects our perception of distance (the so-called "energetics" hypothesis). For instance, Bhalla & Proffitt (1999) found that participants wearing a heavy backpack evaluated the slant of a hill as steeper than participants who did not (assessed using both a visible angle protractor, and a surface that could be haptically matched using the palm to the slant of the hill). However, the distinction between 'perception' and 'cognition' has become especially opaque in the "perception of effort affects perception of slant" literature. Durgin et al. (2009), Shaffer et al. (2013), Firestone (2013), Firestone & Scholl (2016), Firestone & Scholl (2017), and Durgin (2017) all attempt to separate out perceptual and cognitive biases. For instance, Durgin et al. (2009) found that simply providing a cover story for the backpack (such as the suggestion that it was necessary for measuring equipment) negated its effect on slant judgements. Increasingly advocates of the embodied perception approach, such as Schnall (2017), deny that perception and cognition can be separated. But Durgin (2017) is correct to resist this:

> "The … fundamental problem with Schnall's argument is that suggesting that perceptual experience is the same as attribution trivializes the energetics hypothesis. Should we be interested in this work if it is just about judgment? Isn't it the purported effect on perception that made the theory interesting? Conflating (judgmental) attribution effects with perception has been a common artifact of the energetics approach, and this is why many psychologists have rightly lost interest in it. To the extent that energetics theorists have stopped asking Koffka's (1935) question, "Why do things look as they do?", their theories aren't really about perceptual experience at all."

A final criticism of the ground plane as a source of absolute distance information is that it appears to fall into the following vicious circle given other assumptions in the literature:

1. Absolute distance is used to extract slant of the ground plane (from binocular disparity).

But

2. Slant of the ground plane is used to extract absolute distance from the ground plane.

This is a problem for the ground plane account. It would seem that we need an independent source of absolute distance information to scale the slope of the ground plane itself. Nor do I know of any studies that demonstrate (or even attempt to show) that absolute distance from the ground plane can be used to scale binocular disparity, even though this is what one would expect according to most orthodox accounts of depth cue integration.

3. Visual Field Blur: Tilt-shift photography enables us to simulate miniaturisation by applying a linear blur gradient to the scene. For instance, the blur gradient in Figure 8 is consistent with the real-world scene being photographed with a 60 m wide aperture, but this is not the interpretation that we automatically adopt. Instead, we appear to attribute the presence of a linear blur gradient to the optics of the eye itself, where the gradient in Figure 8 is consistent with viewing the scene from 6 cm. Consequently, Held et al. (2010) and Vishwanath & Blaser (2010) suggest that we experience the scene as a miniaturised model. Vishwanath & Blaser (2010) also show that this effect is not reliant upon a recognition of the scale of the scene itself (using non-descript rocks as opposed to a cityscape). However, there are four concerns:

[Removed due to copyright]

Figure 8. Tilt-shift miniaturisation consistent with a real-world scene being viewed with a 60m aperture, that is interpreted as a small scene up close, from Held et al. (2010)

First, interest in absolute size / distance from linear blur gradients arose out of the emergence of tilt-shift photography, and especially simulated tilt-shift photography (so-called 'miniature faking'), on the internet. But just how observable is this blur in everyday real-world viewing? Consider two simulated images from Held et al. (2010):

[Removed due to copyright]

Figure 9. Two examples of simulated tilt-shift miniaturisation from Held et al. (2010). The image on the left simulates a 15cm focal distance, the image on the right simulates a focal plane of 50cm.

The image on the left simulates a 15cm focal distance, the image on the right simulates a focal plane of 50cm. A key difference is the slant of the ground plane. No matter how close you get to a fronto-parallel surface, there won't be any vertical blur gradient in the retinal image. There will be a radial blur gradient, but this is constant across all viewing distances assuming the fronto-parallel surface takes up the same angular field of view. The geometry for fronto-parallel surface taking up 30° of the visual field is the same monocularly for a surface viewed at 20cm as a fronto-parallel surface viewed at 20m. Consider Fig.39 below. This could describe a surface at 20cm or 20m. The % increase in distance of the surface as you move off-axis relative to fixation distance remains the same. So 1 divided by this % difference will remain the same. So the radial blur gradient will remain the same. There's nothing special about radial blur gradients at near distances.

But then the question arises, how likely are we to encounter the required level of defocus blur in our everyday lives. The geometry of the right-hand image is very rare at close distances (less so far distances, where the image was actually taken). By contrast, the viewing angle of the left image is more common at near distances, but note just how close we have to get – 15cm – before visual field blur comes significantly above threshold. And yet, even at 15cm, we're not getting much of a miniaturisation effect.

Second, and where Held et al. (2010) and Vishwanath & Blaser (2010) disagree, is that Held et al. (2010) articulate defocus blur as a triangulation cue. But we have to remember that, much like the ground plane, our estimate of distance is going to be affected by our judgement of the slant of the ground plane in these photos. To explain why, consider that at close distances (i.e. within the range of myopic blur) the increase in blur could be attributed to (a) keeping the slant of the scene fixed and reducing distance or (b) keeping the distance fixed and simply rotating the scene. Now the blur produced will not be identical in both scenarios (because of the way that defocus blur falls off with 1/distance, simply rotating the scene will produce comparatively less blur in the upper visual field). But first, this isn't noticed, otherwise tilt-shift photography (which literally relies on tilting the image plane) and linear blur gradients would not provide the illusion of reduced scale in the first place. And second, as Vishwanath & Blaser (2010) observe, an image with a linear blur gradient "appears substantially closer and appears more slanted."

> "…the perceived slant with blur was larger than in the comparable no-blur condition and biased in the direction of blur. This is consistent with the informal observation that blur gradients make near-frontoparallel surfaces appear more slanted in the blur direction."

It therefore appears that the linear blur gradient is double counted, contrary to Held et al. (2010)'s triangulation hypothesis.

Third, much like the ground plane, the emphasis appears to be on spatial biases – making objects appear closer or more distant – without specifying absolute distance information. Consider Figure 8b above. This image is consistent with a viewing distance of 6cm away. But that clearly isn't the distance attributed. Vishwanath & Blaser (2010) show that in two successive viewings of a sharp image on a display and a blurred image, "the display with the blurred image was set further way, to be matched with a standard no-blur display at 40 cm". How much is indicated in the following sets of results for one representative participant (CH):

[Removed due to copyright]

Figure 10. Results from Experiment 2 in Vishwanath & Blaser (2010) for one representative participant (CH).

The results that interest us are the results on the right (c and d). The black arrow indicates the simulated distance of the image (25cm), and the red arrow indicates how much further back from 40cm the display has to be placed in order to achieve the same "perceived" distance. In the binocular condition, there seems to be very little effect from having visual field blur as a distance cue. And the interesting, and open, question for the monocular condition is whether participants would actually point to e.g. 25cm. I.e. can they extract absolute distance, rather than "nearer" and "further away" than a standard.

Fourth, we return to the question of whether this is a perceptual or cognitive effect. Vishwanath & Blaser (2010) contrast this effect with familiar size:

> "The effect of blur gradients on distance perception suggests that blur-based miniaturization in photography is the result of a visual phenomenon, rather than just a cognitive outcome dependent on relative object size, object familiarity, or familiarity with macro photography."

But there is another possibility, familiarity with myopic defocus blur. As we've already discussed, participants seem to be responding in a very rough sense to the amount of blur in the visual field. And they don't, as Vishwanath & Blaser (2010) show, seem to be that responsive to changes in the geometry of the blur (e.g. shifting it from y axis blur gradient to x-axis blur). So the participants don't need familiarity with macro photography, but they may well have familiarity with the optics of their own eye, specifically: "I only experience this kind of blur when I'm up close."

Vishwanath & Blaser (2010) suggest:

> "The various scaling effects in pictures supports a view that cognitive sources of information like familiar size likely do not operate as normal visual cues, in the sense that they do not combine with or scale other distance cues (Gogel, 1969; Predebon, 1993). Rather, they might only operate to disambiguate perceptual judgments in a categorical manner under high levels of sensory uncertainty, as is present in pictures, and be overridden to conform to even relatively high threshold optical signals such as blur or accommodation."

But recall that Gogel (1976) used motion parallax as a test for perceptual vs cognitive effects, finding that familiar objects do not have the perceived motion parallax that one would predict from subject head motion had familiar size determined their perceived distance, and therefore concluding that familiar size is merely cognitive. The same criterion can be applied to visual field blur. Would we expect an image displayed with visual field blur to have different motion parallax dynamics than an image displayed without? There seems to be little evidence to suggest this.

4. Motion Parallax: Motion parallax has proven a largely ineffective size and distance cue in virtual reality (Beall, Loomis, Philbeck, & Fikes, 1995; Luo, Kenyon, Kamper, Sandin, & DeFanti, 2007; Jones, Swan, Singh, Kolstad, & Ellis, 2008; Jones, Swan, Singh, & Ellis, 2011; Luo, Kenyon, Kamper, Sandin, & DeFanti, 2015), leading Renner, Velichkovsky, & Helmert (2013) to conclude that "there is no empirical evidence that providing motion parallax improves distance perception in virtual environments." It is for similar reasons that Thompson et al. (2011) and Creem-Regehr, Stefanucci, & Thompson (2015) leave a '?' next to motion parallax as an absolute distance cue. Similarly, Rogers (2019) regards motion parallax as a merely relative depth cue.

It's often thought that Combe & Wexler (2010) is a powerful counterexample. They note that: "It is commonly assumed that size constancy - invariance of perceived size of objects as they change retinal size because of changes in distance - depends solely on retinal stimulation and vergence, but on no other action-related signals." They challenge this conclusion by demonstrating that motion of the observer can also influence size constancy. Combe & Wexler (2010) asked participants to judge whether a displayed target was smaller or larger than the average (Mckee & Welch, 1992; Morgan et al., 2000), and between trials changed the participant's distance from the display either by (a) moving the observer closer/further away, or (b) moving the display closer/further away. This was performed in darkness (with the participant fixating on a point of light a constant distance from them), so that when the target was presented, the retinal image was the same in both conditions. Participants showed better size constancy when the participant actively moved towards the stimulus (active condition) than when the stimulus was moved towards the participant (passive condition). The results from this experiment are shown in Figure 11:

[Removed due to copyright]

Figure 11. Results from Experiment 1 in Combe & Wexler (2010) for three sets of movements, 57.3cm +/- 5cm, 10cm, or 15cm. SM = subject movement, OM = object movement.

What Figure 11 shows is that size constancy is approximately 67% better when it relies on motion from the observer in the z-axis than when it relies on binocular disparity / vergence alone.

In one sense that isn't that surprising. What is required for size constancy is an estimate of the change in distance in the z-axis, and now we have two distinct cues – observer motion and binocular cues – to estimate that change in distance. Although it would be interesting to see how effective size constancy is when participant motion is combined with monocular viewing (reducing the size constancy cue back down to one; unless you think of accommodation as an effective absolute distance cue, which Mon-Williams & Tresilian, 2000 reject).

Combe & Wexler (2010)'s results are similar to Gregory & Ross (1964). Whilst Combe & Wexler (2010) tested size constancy with the observer motion occurring between trials, Gregory & Ross (1964) tested size constancy during observer motion. Participants were moved in a swing (swing amplitude 76cm) as they looked at a circle on a display 1.61m away (+/-38cm away) (see Fig.12). During the swing the angular size of the target was changed until it appeared to be a constant visual size during the swing. The participants were tested monocularly, thereby beginning to answer the question I just posed with regards to Combe & Wexler (2010), namely whether this effect would be present monocularly? Gregory & Ross (1964) found a monocular size constancy effect of 0.14-0.18 (where 0 is no size constancy, and 1 is full size constancy), suggesting a small effect of passive bodily motion on perceived size.

[Removed due to copyright]

Figure 12. Apparatus in Gregory & Ross (1964)

One question about these kinds of findings, however, is that they seem to be reliant on already knowing the absolute distance of the object in question. For instance, this perceptual expansion does not occur when we approach a distant object (e.g. objects on the horizon) in a car. As we drive down a highway at speed, objects on the horizon don't seem to expand, even though they keep a constant retinal image, and this would be predicted by Gregory & Ross (1964)'s results. To illustrate the point with using Gregory & Ross (1964)'s stimulus (which is a disc taking up 0.57° of the visual field), the moon is a 0.5° disc suspended in the sky. If we drive down a highway monocularly at night, why should we not expect the moon to similarly shrink in apparent size in the same way that the constant angular size stimuli did in Gregory & Ross (1964)? If the answer is that it is because we know the moon is far away, then this suggests that this effect is highly reliant on prior knowledge about distance. Therefore Gregory & Ross (1964)'s finding appears to be heavily reliant upon prior knowledge about the experimental apparatus and stimulus (at least enough to tell us that what we are looking at is an object up close rather than on the horizon).

Indeed, we can push this thought even further. Imagine a participant sitting on the front of a constantly moving train in darkness. Here we no longer have the vestibular feedback telling us about passive motion, since vestibular feedback is reliant on acceleration. Now we optically present a monocular target to them of a fixed angular size. Does this target appear to shrink as predicted by Gregory & Ross (1964)'s results? On the one hand, it would be weird if the effect was only present during acceleration not constant motion. On the other hand, if we did find an effect, what would be the source of the motion information apart from the participant's own subjective knowledge about the fact they are moving at a constant speed?

What these discussions point to is that this size constancy effect appears to be merely cognitive in nature, and certainly heavily reliant upon subjective knowledge. I will address two further concerns. First, that on closer inspection the results from Combe & Wexler (2010) appear to be ambiguous. Second, that whatever our position on Combe & Wexler (2010) and Gregory & Ross (1964), this still doesn't tell us anything about motion parallax.

The problem with Combe & Wexler (2010)'s results, by their own admission, is that their black-out conditions were not perfect. They worried that perhaps "subjects faintly perceived the edges of the stationary monitor", which benefitted the 'participant motion' condition (where this provides a looming cue), vs. in the 'stimulus motion' condition where the display was fixed, and the motion of the stimulus came from changes in binocular disparity off the display plane (so no looming from display edge). So Combe & Wexler (2010) tried a new condition where the display moved towards the observer for the 'stimulus motion' condition, thereby truly matching the retinal images in the two conditions, with the only difference being the passive observer motion. But there are two things notable about their results (Fig.13). First, the SM (subject motion) condition leads to completely different results (close to perfect size constancy), with a close to 60% improvement in size constancy, despite this SM condition being identical to their previous SM condition (it was only the OM condition they changed). There is no explanation for this fact. Second, the OM (object motion) condition now shows over-constancy compared to the SM (subject motion) condition. Combe & Wexler (2010) argue that because the SM condition shows less over-constancy, this is consistent with their account that vestibular cues contribute to size constancy. But the fact that there is more variability between repeated tests of their OM condition than between their OM and SM conditions leaves these results in an ambiguous state.

[Removed due to copyright]

Figure 13. Results from Experiment 3 in Combe & Wexler (2010) when OM = display moving towards the observer, rather than changing disparity.

Finally, we have to be clear that what was being demonstrated in Combe & Wexler (2010) is not motion parallax. If the ground plane is the ability to scale the geometry of the scene using the height of the observer in the y-axis, motion parallax is the ability to scale the scene using the lateral motion of the observer in the x-axis. Assuming the eye is fixed in a neutral position (and this is an assumption that accounts implicitly assume without recognising), then the closer an object is, the more the object will move with lateral motion by the observer in the x-axis.

The classic discussion of motion parallax as an absolute distance cue is Gogel & Tietz (1973), but other recent accounts that express a commitment to "absolute motion parallax" include Beall et al. (1995); Glennerster et al. (2006); Ooi & He (2015); de la Malla et al. (2016).

There's no reason in theory why it couldn't also apply to the motion of the observer vertically in the y-axis. Rather than swaying side to side, an observer could simply sit then stand up. The closer an object is, the more it should move, assuming eyes fixed in neutral position. I know of no experiments which have tested this, but if the visual system is able to triangulate absolute distance from lateral body motion, then there's no immediate reason why vertical bodily motion shouldn't be just as effective. In the case of y-axis motion, motion parallax would be analogous to the ground plane, using the known height of the observer to scale the scene. In the case of x-axis motion, motion parallax is analogous to vergence, but instead of triangulating between two eyes simultaneously, we triangulate one (or two) eyes over time, with the motion of the observer in the x-axis taking the place of IPD. So both vergence and motion parallax rely on multiple perspectives in the x-axis, the difference is that in one case these multiple perspectives are simultaneous (vergence), whilst in the other they're sequential (motion parallax).

The claim of motion parallax is that the visual system is able to use x-axis or y-axis motion to extract z-axis distance from the retinal image. By contrast, experiments such as Combe & Wexler (2010) have no role for the visual system in extracting z-axis absolute distance from the retinal image. Instead, z-axis absolute distance is directly provided by proprioception (the sensation of motion). So the visual system isn't computing z-axis distance.

So far as motion parallax is concerned, there is little evidence that x-axis or y-axis motion actually improves absolute size and distance judgements. In terms of the visual system translating x-axis motion into z-axis motion, Renner et al. (2013) find that "there is no empirical evidence that providing motion parallax improves distance perception in virtual environments." It worth pausing to consider why this might be? There are three reasons:

1. First, the motion parallax claim is often driven by a suggestion that it is more ‘ecologically valid’ than the static presentation of a visual scene. But the actual movements being studied – swaying laterally as indicated by the following video – have little ecological validity. We rarely adopt such motions, so this can’t be how we judge distances in everyday viewing.

[Removed due to copyright]

Figure 14. Participants swaying left to right in Svarverud et al. (2012). Link: https://www.glennersterlab.com/Videos/DistTrans.mp4

More recently interest has focused on ‘micro motion parallax’, the more ecologically valid small movements we all make when sitting or standing. But de la Malla et al. (2016) recently found that micro motion parallax (a) had virtually no effect (monocularly or binocularly) on the distance estimates of an isolated object, (b) had virtually no effect, even with the introduction of relative motion parallax cues, on binocular distance estimates, and (c) had only a negligible effect (gain of 0.115) on monocular distance estimates with the introduction of relative motion parallax cues.

2. Second, another crucial challenge to the motion parallax account is the fact that the visual system works to eradicate the motion parallax signal. The basis of absolute motion parallax is that near objects move on the retina more than distant objects. But this is only if the eyes are held in a fixed position. Instead, what actually happens is that our eyes fix on an object of interest and rotate to maintain fixation upon the object of interest. The key point being that in order to maintain visual stability the visual system eradicates the information required for motion parallax.

An extreme case is found in birds, where their head remains fixed in space as their body moves (Frost, 2009). What use would bodily motion signals be to a bird trying to extract absolute distance from motion parallax if visual stability from their neck muscles ensured that the retinal image remained fixed no matter what their bodily movements? The same principle applies to humans. The key difference between birds and humans is that humans have moveable eyes that rotate in their eye socket, meaning that rather than the visual field remaining fixed with neck movements (perhaps explaining why many birds have long necks), for humans it remains fixed with eye movements (although neck movements are also an important part of human visual stability: think about fans sitting alongside a NASCAR track, and watching the cars streak past from left to right; the fans will follow the cars with head rotations, and not just eye rotations). The dilemma facing the absolute motion parallax account is that the information that we need to preserve in order to extract absolute motion parallax, is the very same information that we need to eradicate in order to maintain retinal stability.

A good illustration of this is motion parallax for a single point in space. In theory motion parallax should provide this distance as a simple triangulation cue. If the monocular eye is fixed (let us assume looking forward), then as the observer moves laterally (side to side) in the x-axis, a closer object will move more on the retina than an object further away. The reason for this is the reduction in angular size with distance due to perspective. As you move left by 1m in the x-axis relative to the world, the world moves right by 1m in the x-axis relative to you. But what does this 1m translation look like in the retinal image? A translation by 1m viewed from 2m away is going to produce very significant motion on the retina. By contrast a translation by 1m viewed from 200m away is going to produce virtually no motion on the retina. And this is how computer vision systems can take advantage of motion parallax. But in computer vision systems the camera remains fixed. By contrast, if the human eye tracks a point of light as the observer moves laterally (side to side) in the x-axis, there should be no motion on the retina, and so no retinal cues to the distance of the point of light, only the oculomotor and head movement cues as the eyes and the head rotate to maintain retinal stability. So absolute motion parallax becomes essentially a version of vergence, but unlike vergence it relies on:

1. Estimating the motion of the observer, rather than relying on a fixed IPD which could be learnt over time, and the visual system could calibrate itself to.

2. Estimating the change in eye rotation over time, rather than instantaneously via vergence.

The fact that the visual system would have to extract eye rotation over time is something that is discussed (and seemingly assumed) by Nawrot et al. (2014). One argument for assuming this can be achieved is that the visual system needs to know distance in order to track the stimulus accurately. When the observer moves left in the x-axis, then the world will move right in the x-axis relative to the observer. If the eye is to maintain retinal stability, i.e. eradicate this rightward motion in the x-axis, then the only thing it can do is rotate to keep whichever object is being fixated on fixed on the fovea. If the object is far away, this will require a minor eye rotation rightwards. But if the object is close, this will require a major eye rotation. So it seems that in order to maintain retinal stability, the visual system already needs to know the distance of the fixated object.

Gogel & Tietz (1973) argued that we could use this fact of retinal stabilisation to get at the absolute distance of the object encoded by the visual system. Gogel & Tietz (1973) found that points of light in darkness were most effectively stabilised around 2-3m (the so-called 'specific distance tendency'). That seems quite plausible. Essentially what it tells us that these retinal stabilisation mechanisms have an effective range of motion on the retina. But it is huge jump from this to the suggestion that for distance perception, in reduced cue environments, objects are encoded as being seen as 2-3m away, which was Gogel & Tietz (1973)'s claim.

3. Third, we know motion parallax isn't necessary. 1. As we've already discussed we rarely adopt the swaying lateral motion required. 2. If we're introduced into a new scene (via virtual

reality or by having a blindfold removed) we can gauge the gist of the scene's scale pretty instantaneously, before time for anything more than micro-motion parallax has elapsed. 3. Nor does our ability to judge distances (when introduced into a scene via VR) appear to be especially compromised by sitting down and reducing bodily motion. 4. As we shall see in Experiment 1 and Experiment 2, participants are pretty accurate in pointing with a hidden hand in full-cue conditions when their head movement is restricted by a chin rest.

5. Vertical Disparities: Vertical disparities are the differences in the height of points in the visual field in the two eyes because (a) y-axis angular height falls off linearly with distance, and (b) apart for points on the meridian, the addition of the interpupillary distance (IPD) increases the distance to the point in one eye relative to the other. This was hypothesised as a basis for binocular distance perception by Longuet-Higgins (1981), Mayhew & Longuet-Higgins (1982), and Gillam et al. (1988), but was found to be ineffective as an absolute distance cue by both Cumming et al. (1991) and Sobel & Collett (1991). Rogers & Bradshaw (1995) hypothesised that the reason Cumming et al. (1991) and Sobel & Collett (1991) didn't find an effect is because of the their small visual angle stimulus, explaining why they found an effect with a 80° x 80° stimulus at 57cm. In the Appendix A of Rogers & Bradshaw (1995), vergence was fixed at 57cm, and participants were asked to verbally estimate in cm the absolute distance of a surface defined by vertical disparities between 28.5cm and infinity. Rogers & Bradshaw (1995) report distance scaling "equivalent to 65% of that required for perfect scaling", but (a) distance in this analysis is measured in angular terms as if it is vergence distance (so 65% scaling means a surface at infinity is judged to be about 2m away), and (b) this analysis is achieved after heavy normalisation of the distance estimates:

> "Quantitatively, different observers differed in the way they used numbers to indicate absolute distances and so the distance estimates were normalised by expressing each subject's six distance estimates as z-scores and then reconverting them back to distances by using the average distance and average standard deviation of the six subjects."

One is therefore left wondering whether participants achieve anything close metric distance estimates, or indeed anything more than ordinal distance estimates, if we remove this kind of averaging over all participants (recall a similar critique of Fisher & Ciuffreda, 1988 by Mon-Williams & Tresilian, 2000 in the context of accommodation).

Rogers & Bradshaw (1995) are right that vertical disparities are maximised for large field of view surfaces. On the face of it, this might seem counterintuitive. Consider the following example in Fig.15. Where will the vertical projection of points on that surface be at their greatest difference between the two eyes? I.e. where are the differences between α and β going to be maximised for a fronto-parallel surface?

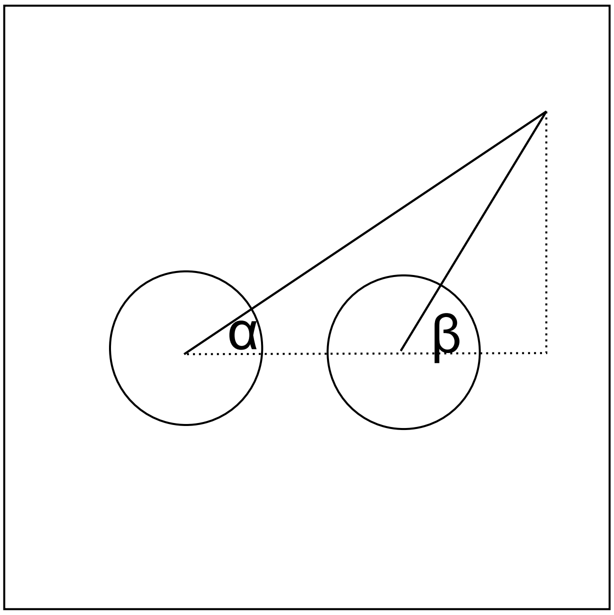

Figure 15. Vertical disparity is the difference between α and β.

Angles α and β are a function of three distances analysed in terms of the x-axis, y-axis, and z-axis relative to the observer's head (x-axis runs though eyes): First, the y-axis height of the point on the fronto-parallel surface. But this is the same for both eyes. So what matters then is the distance of the point from each eye when the scene is viewed from above (Fig.16). Second, this will be affected by the distance in the z-axis of the surface from the eyes. But again, the z-axis distance of the point is the same for both. So the only difference that can differentiate α and β is the x-axis position of the point. So when will the difference between α and β be maximised? On the face of it, you might think this is whether the point is directly in front of one eye, so the x-axis distance is reduced to 0 for one eye, and 6.2cm (IPD) for the other, for instance point A in Figure 16 (note that disparities are rotationally invariant, so it doesn't matter where the eyes are actually looking, both eyes could be parallel).

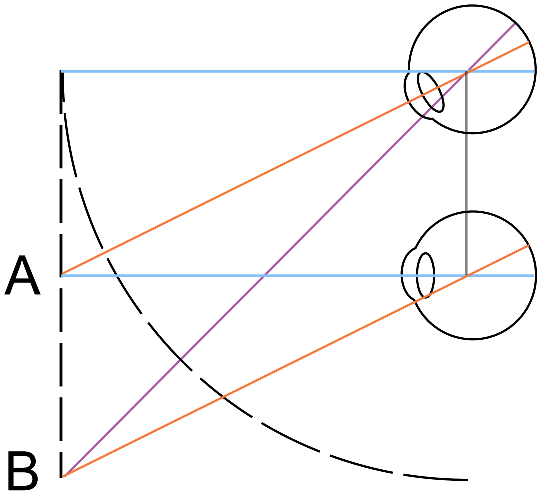

Figure 16. Calculating the distance of points on a fronto-parallel surface to the eye

However, the reason why this is a mistake is easy to appreciate. Think about the right-angle triangle formed between A and the two nodal points of the eyes. Obviously, if we move rightwards from A (upwards in this diagram), this triangle will become an equilateral triangle on the meridian, as the distance to the right eye reduces, and the distance to the left eye increases. By contrast, if we move leftwards from A to B (downwards in the diagram), the right angle of the triangle becomes more and more obtuse, so the ratio of 'distance from the right eye' to 'distance from the left eye' increases. And, if we were to continue left, the angle will always get more and more obtuse. You might therefore think that the ratio between the two distances will always increase. But at some point, this effect which is due to the IPD of 6.2cm gets drowned out by the sheer magnitude of the distances. We can illustrate this using a heatmap. Instead of Rogers & Bradshaw (1995)'s 80°

x 80° stimulus, we will use a 800° x 800° stimulus at 57cm to illustrate the full effect, even though the human visual system limited to 120° x 120°:

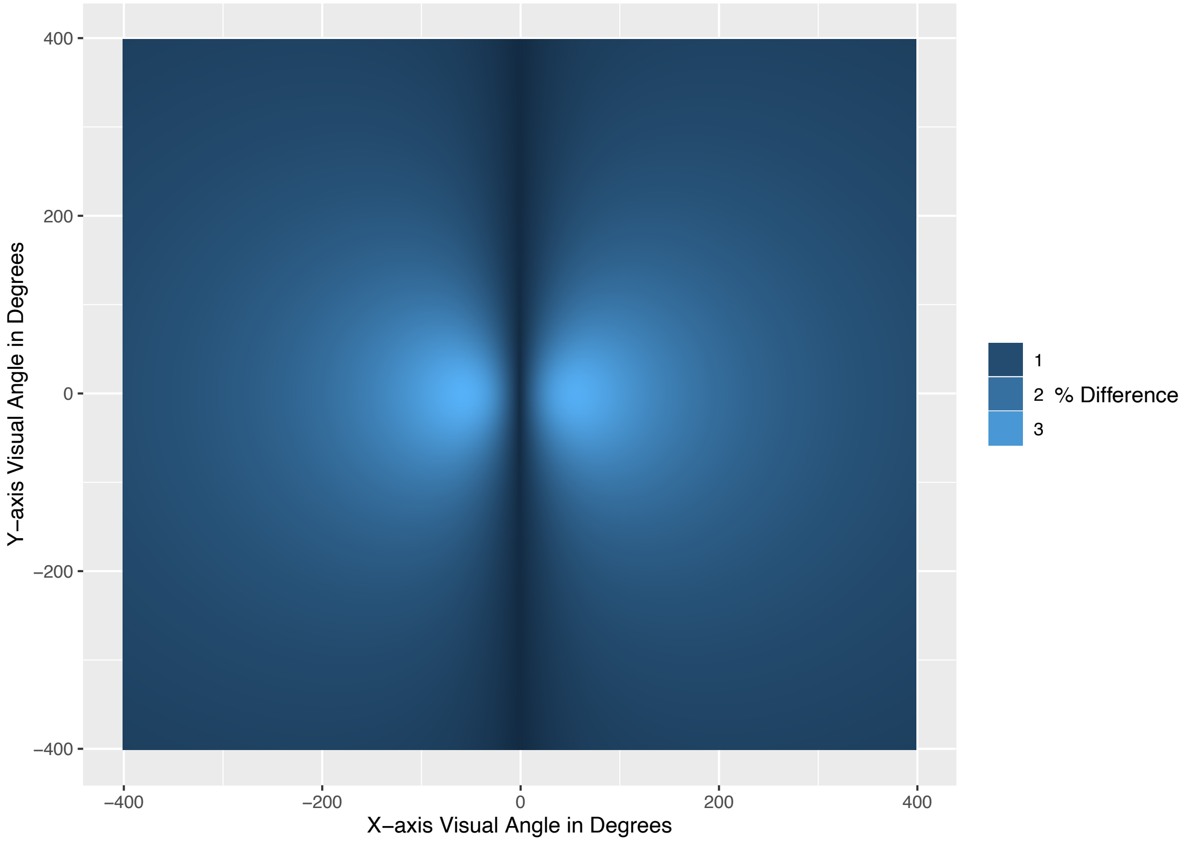

Figure 17. Percentage difference in angular height of points between the two eyes for a 800° x 800° stimulus viewed at 57cm.

Vertical disparities are maximised in % terms around eye level, because by minimising the y-axis height of the point, you maximise the x-axis and z-axis effect described in Fig.16 on the total distance to the eye. But what matters to us isn't the percentage difference between the two angles α and β, but the absolute difference between these angles. First, whilst human subjects in experimental conditions may be influenced by Weber's Law on just noticeable differences in terms of stimulus size, this should not affect retinal disparity detection for individual points on the stimulus. Second, whilst the acuity of the eye may degrade the perception of angles at extreme eccentricities, disparities are rotationally invariant, so could be picked up with eye movements to the periphery. When we replot Fig.17 in absolute terms, we get a quite different distribution:

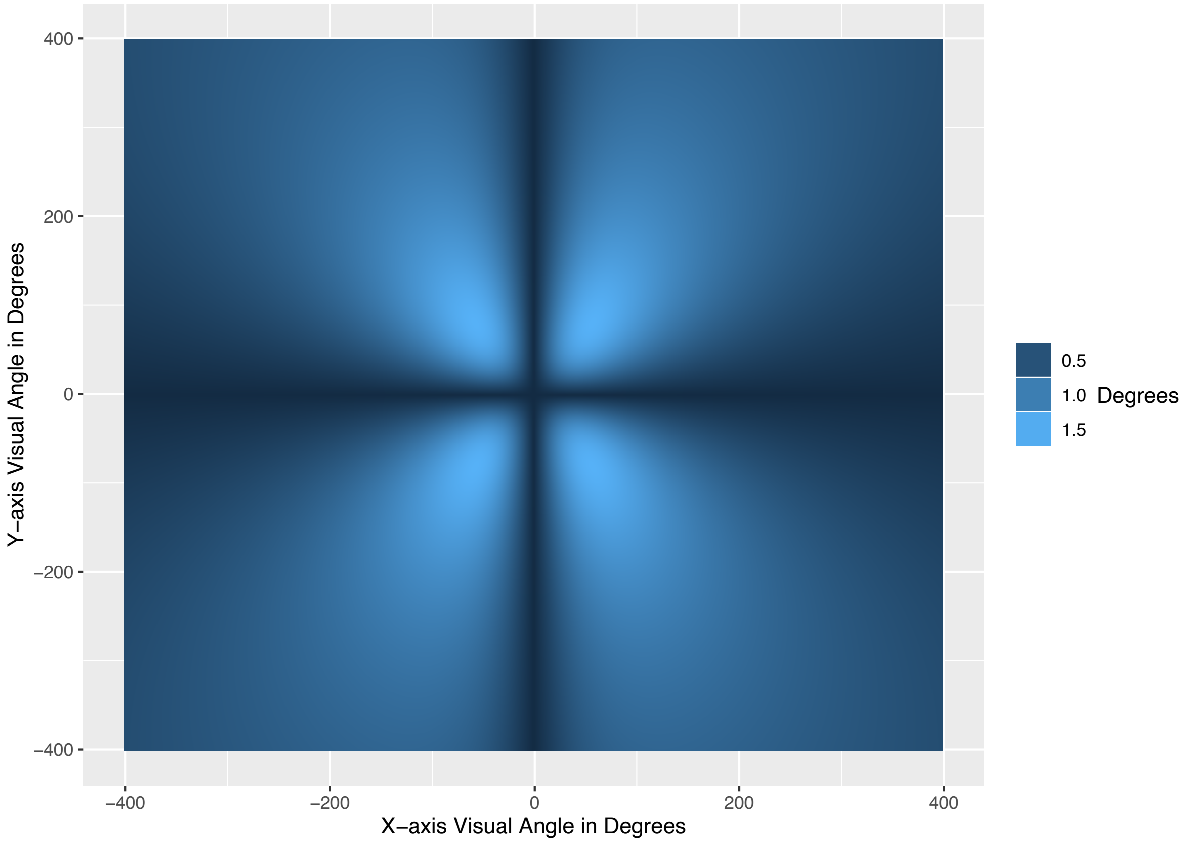

Figure. 18. Absolute difference in angular height of points between the two eyes for a 800° x 800° stimulus viewed at 57cm.

Taking the central 80° x 80° portion of this 800° x 800° stimulus to represent the actual stimulus in Rogers & Bradshaw (1995):

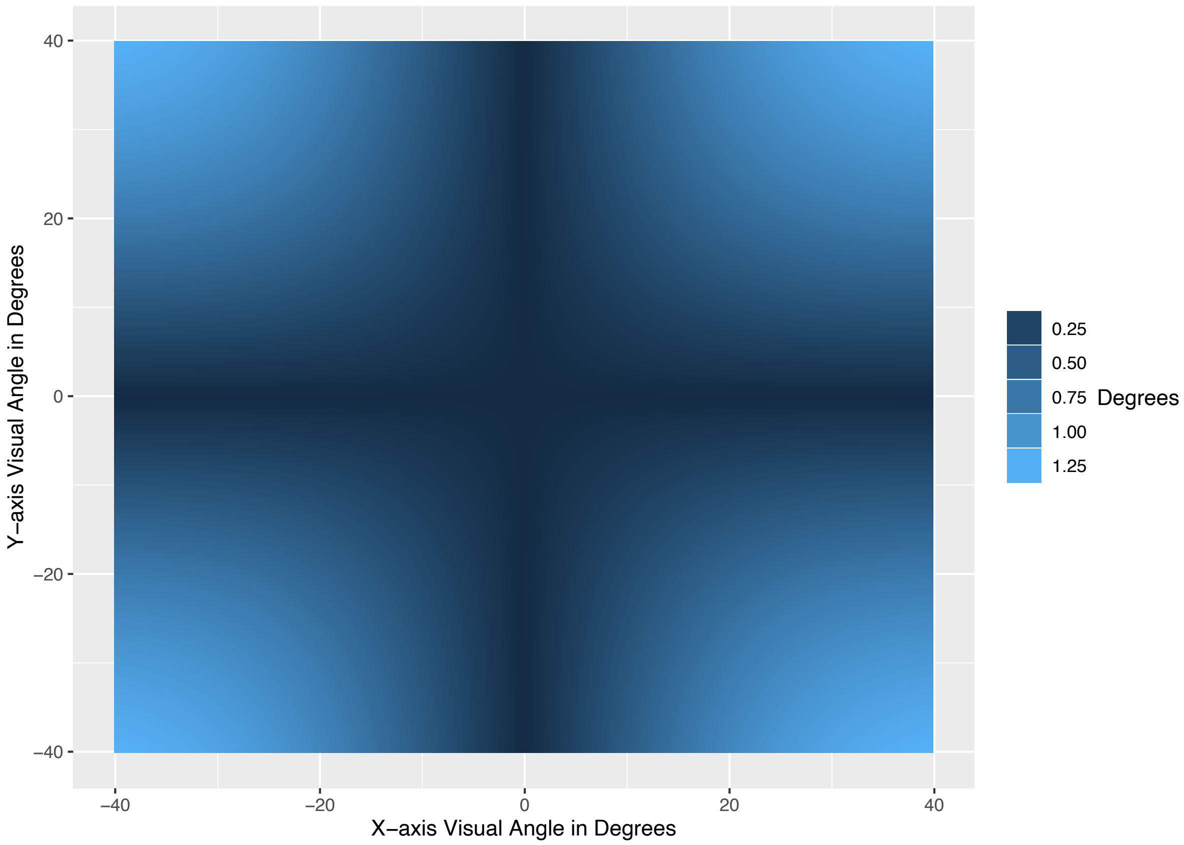

Figure 19. Absolute difference in angular height of points between the two eyes for a 80° x 80° stimulus viewed at 57cm.

So Rogers & Bradshaw (1995) are right to emphasise the extremities of their 80° x 80° stimulus. But there are two concerns about Rogers & Bradshaw (1995):

First, we rarely encounter objects that take up 80° x 80° of the visual field. It seems problematic that the stimuli in Sobel & Collett (1991) subtended 25° x 30° and they found no effect, even at distances as close as 12.5cm. So how useful is a cue that only works once the object in question takes up something close 80° x 80° of the visual field? It's been suggested to me that the horizon is one such object, but recall there are no vertical disparities on the horizon (Fig.18).

Second, Rogers & Bradshaw (1995), and the literature since Rogers & Bradshaw (1995), think about vertical disparities in terms of the angular rotation of the stimulus, and this is a mistake. The key articulation of this principle is Fig.6 of Rogers & Bradshaw (1995).

[Removed due to copyright]

Figure 20. Figure 6 from Rogers & Bradshaw (1995)

There are two rotations of the stimulus in Rogers & Bradshaw (1995). First, the desired rotation of the stimulus, i.e. what the retinal image should look like to the observer, which is captured by the bottom row of Fig.20 (Fig.20, d-f), and second, the required image to put on a screen at 57cm to achieve these retinal images, captured by the top row of Fig.20 (Fig.20, a-c).

First, the rotation of the stimulus in the top row is a mistake. Take Fig.20(c). You can't eradicate an oblique view of a picture by pre-rotating the picture in the opposite direction. That is obvious if you think of a more extreme example. Imagine the crowds are so big at the Louvre that many people can only view the Mona Lisa from the side, say 70° off centre. Imagine someone views the Mona Lisa from the right. The problem is that the Mona Lisa will appear rotated 70° clockwise. Now imagine the museum authorities want to give these visitors a better impression of what it is like to view the Mona Lisa front on by placing an image of the Mona Lisa rotated counter-clockwise by 70°, to be viewed clockwise at a 70° angle by the observer, in the hope of cancelling these rotations out and presenting a fronto-parallel Mona Lisa to the observer. What the viewer will experience won't be a frontal view of the Mona Lisa. Instead, it will be a distorted view of an already distorted image. It's true that to present a frontal view of an image on a rotated surface you'll have to present a pre-distorted image, but that won't be a simple rotation, as the anamorphic projection in Holbein's 'The Ambassadors' (1533), or OpenGL in Fig.30, below both illustrate.

Second, the rotation of the stimulus in the bottom row of Fig.20 is also a mistake. That these are not the correct retinal images can be seen if we consider Fig.20(d). This is supposed to simulate the retinal image of a fronto-parallel surface at 28.5cm. But the key problem is that binocular disparities are rotationally invariant. How the eyes rotate doesn't affect the light rays converging to the nodal point of each eye since the nodal point is roughly at the centre of rotation. So Fig.20(d) also has to be an accurate representation of the retinal images when the two eyes are fronto-parallel looking at the screen 28.5cm away. But this is exactly not what we would experience. Instead, we would experience are two fronto-parallel surfaces, like 20(f).

What is the cause of the confusion here? It focuses on the fact that whilst the human eye is rotationally invariant, cameras are not. So it is a mistake to approximate the human retinal image Fig.20(d-f) as if it is a projection onto a fronto-parallel surface (as it the case with a pinhole camera). That's one key difference between the eye and a pinhole camera: there is no rotation of the image with eye rotations. All you are doing by changing the distance of a stimulus with a fixed angular size is shifting the image left or right on the retina. We can see this by using OpenGL to render a stimulus 57cm away to be viewed on a display 57cm. It is an anaglyph (red = left eye, blue = right eye), but because the rendered distance (57cm) is the same as the display (57cm), no binocular disparities are simulated in the image.

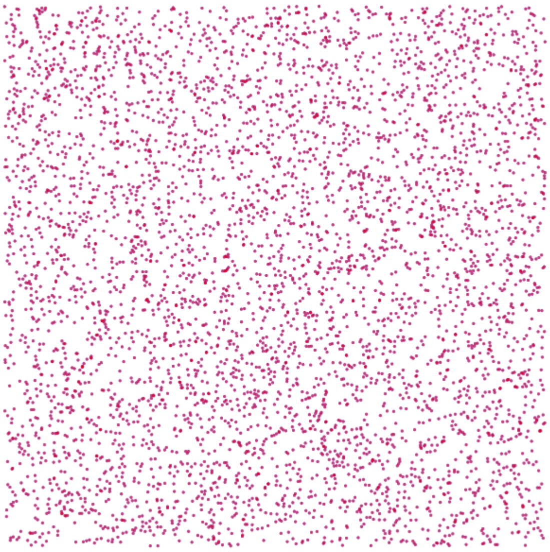

Figure 21. OpenGL simulation of a stimulus 57cm away to be displayed on a display 57cm away. Because these two distances coincide, this is simply a 2D stimulus.

Now what happens when we double the distance of the right eye (blue) stimulus to 114cm, but also double the size of the stimulus to 160cm x 160cm?

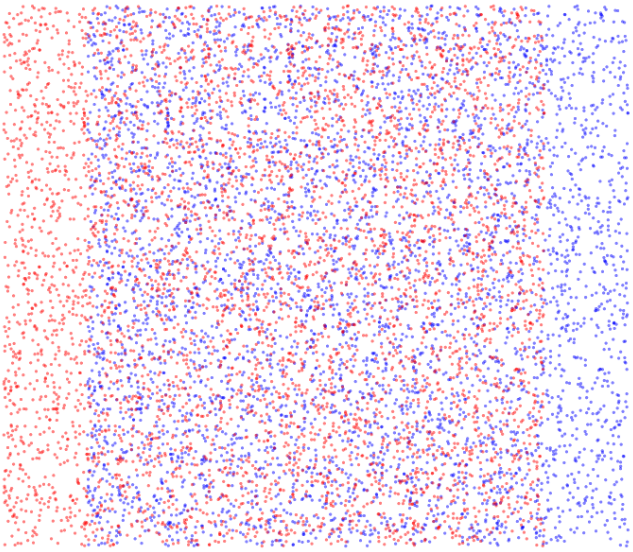

Figure 22. Figure 21, but with the blue right eye stimulus now representing a stimulus at double the distance and double the height and width.

As Fig.22 demonstrates, all you are doing when you change the distance of the stimulus, is to shift the stimuli horizontally relative to each other by roughly the angular size of the IPD at the distance of the stimulus (atan(IPD/distance)). Physically what this means is that the visual scene still projects to exactly the same location in physical space. All that changes is the position of the fovea in physical space. But as you can see from Fig.22, there are no vertical disparities so far. So should we reject the very notion of vertical disparities? This where the second disanalogy between the eye and a pinhole camera comes in. Unlike a pinhole camera, because the retina is roughly spherical, we experience a barrel distortion when looking at a fronto-parallel surface.

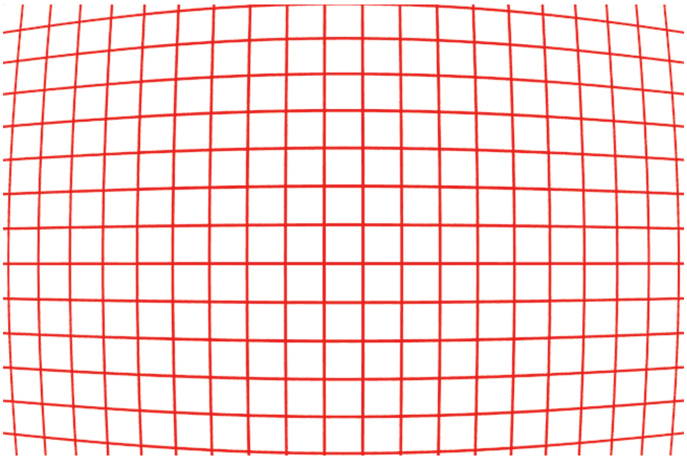

Figure 23. Example of barrel distortion from https://photographylife.com/wp-content/uploads/2013/08/Barrel-Distortion.png. Note that this is not a geometric rendering of the retinal image.

Note that this barrel distortion is not distance dependent, we just don't ordinarily notice it at far distances because fronto-parallel surfaces rarely take up a sufficiently large amount of the visual field to notice it. Nor do we particularly notice them up close. For instance, the grid in Fig.20(f) is subject to them when we look at it up close on a page, and yet the grid looks undistorted. But what changing distance does do, in binocular vision, is to shift these barrel distortions laterally relative to one another in much the same way that the stimuli were shifted in Fig.22. So the resulting retinal image will look something like this:

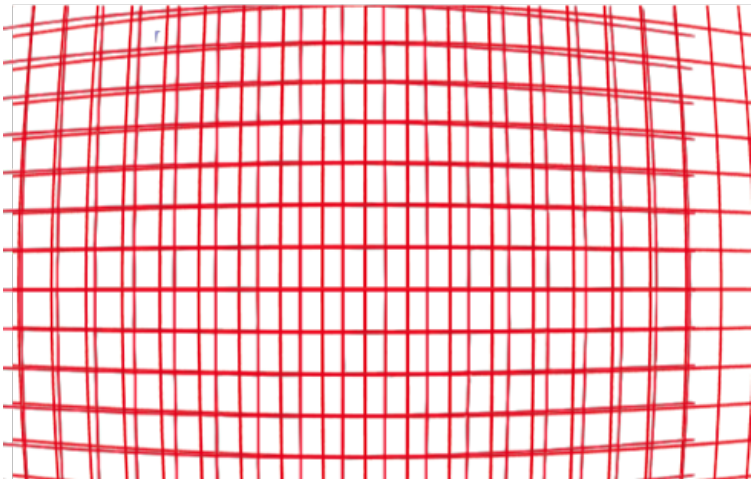

Figure 24. Simulation of vertical disparities as two barrel distortions shifted horizontally with regards to one another.

This will produce changes in the individual heights of points, as discussed in the context of Figs. 17-19, but they won't be apparent in anything like the same way that Rogers & Bradshaw (1995) hypothesise because there isn't any rotation of the stimulus as there would be if the eye was a camera. The key point being that there are no keystone distortions of the retinal image (the stimulus in Rogers & Bradshaw, 1995), only barrel distortions. By contrast, we know that the vertical disparities were properly presented in Cumming et al. (1991) and Sobel & Collett (1991) because the stimuli were physically moved in distance rather than rendered. But in neither of these studies were vertical disparities found to be effective absolute distance cues.

This conclusion about distance perception is consistent with related discussions in shape perception, where the question is whether the visual system is able to extract absolute values from binocular disparity that are invariant to vergence changes. As Lunn & Morgan (1997) note, changes in disparity curvature are roughly invariant to distance changes. So, if the human visual system was evolved to extract absolute properties from the environment, we would expect observers to be sensitive to changes in disparity curvature. But Lunn & Morgan (1997) found that Weber fractions were 4%-10% for relative disparity changes, 6%-12% for changes in disparity gradient, and 15%-30% for curvature. They therefore conclude that their study "supports the view that human stereoscopic vision aims to represent the local scene relative to the observer, at the expense of computing intrinsic properties of objects, such as curvature."

6. Accommodation: Accommodation is the focal power of the eye. Immediately, this doesn't strike us as a likely absolute distance cue because accommodation drops off with age (presbyopia) meaning that it is largely ineffective for most of the population over 40. However, evidence in favour of accommodation as an absolute distance cue comes from Fisher & Ciuffreda (1988), who had young subjects point to the distance of a target viewed in darkness in a Badal system (to control for size and luminence), and found a relationship of $y = 0.27x + 2.33$ for distances between 16cm and 50cm. This low average gain of 0.27 masked a high degree of individual and inter-stimulus differences: individual gains ranged from 0 to 0.8 for a cross stimulus, and –0.20 to 0.92 for a Snellen chart. Mon-Williams & Tresilian (2000) found similar results (varying stimulus size between trials, rather than using a Badal system), but unlike Fisher & Ciuffreda (1988) concluded that accommodation is not an absolute distance cue because even when they focused on the two observers with the highest gains ($y = 0.62x + 1.55$ and $y = 0.60x + 0.88$ for 16cm-40cm), they argued that the variability in their responses was so great that "it is clear that accommodation is providing no functionally useful metric distance information for these observers. The responses were unrelated to the actual distance of the target."

Mon-Williams & Tresilian (2000) instead suggest that subjects may only be able to extract ordinal depth information from accommodation (i.e. whether the present trial is closer or further away than the previous one). But this doesn't explain how two subjects could have the same ordinal depth success rate (78-79%) but completely different gains ($y = 0.62x + 1.55$ vs. $y = 0.28x + 1.49$) in their experiment, nor the fact that when Liu, Hua, & Cheng (2010) presented subjects with a monocularly viewed half Siemens star, and asked participants to judge the change between three possible accommodative distances (20cm, 33cm, 50cm), the average success rate was 80% (with individual performance ranging from 67% to 93%), even though merely ordinal depth would leave participants at chance on 2 out of the 3 trials (if the starting value was 20cm, merely ordinal depth couldn't distinguish 33cm vs 50cm, and if the starting value was 50cm, merely ordinal depth couldn't distinguish 20cm vs 33cm).

So accommodation may be a more effective cue than Mon-Williams & Tresilian (2000) suppose. However, Fisher & Ciuffreda (1988), Mon-Williams & Tresilian (2000), and Liu et al. (2010) all rely on blur driven accommodation. But this introduces three confounding retinal cues: first, the initial amount of blur in the stimulus, and second, the changing retinal blur as the stimulus comes into focus, and participants might strategize to achieve a gain of 0.27 based on a rough heuristic of more blur = closer. Another concern is that the sudden presentation of the stimulus, as close as 16cm in Fisher & Ciuffreda (1988), is going to put a lot of strain on the eyes, and participants might simply be responding to this, rather than any actual visual processing.

A better, but more phenomenological, test of accommodation as an absolute distance cue is to move an image slowly behind a Badal lens. This would eradicate the sudden change in defocus blur in previous experiments, whilst controling for angular size and luminance. I report in Linton (2017), p.132 the phenomenological observation that moving a target back and forth behind a Badal lens appears to have no effect on its perceived distance. But how can we operationalise this observation, so we're not just relying on introspection? The problem is that Badal setups prove notoriously difficult for most subjects (see Charman & Heron, 2015), with only 16% of the subjects in Metlapally, Tong, Tahir, & Schor (2014), and 20% of the subjects in Metlapally, Tong, Tahir, & Schor (2016), able to accommodate effectively. So, rather than trying to isolate accommodation as an individual cue, a more effective approach (which I adopt in Experiment 2) is to incorporate accommodation alongside vergence, and test their effectiveness in combination.

3. Vergence as an Absolute Distance Cue

The cues we have discussed so far are either (a) largely ineffective (motion parallax, vertical disparities, accommodation), (b) limited to very specific contexts (ground plane, visual field blur), or (c) merely cognitive in nature (familiar size). So the only potential visual cue to scale of general application remaining is vergence, the angular rotation of the eyes. This is why vergence is the focus of the three experiments in this thesis.

Since Kepler (1604) and Descartes (1637), vergence has been regarded as one of our most accurate and reliable absolute distance cues at near distances, where such accuracy is paramount for reaching and grasping. This is for eight reasons:

1. Triangulation: Extracting absolute distance from vergence relies on simple principles of geometry. There is no need to infer 3D content from the 2D retinal images. Instead, the visual system is able to triangulate distance from the rotation of the eyes (Parker, Smith, & Krug, 2016; Banks, Hoffman, Kim, & Wetzstein, 2016; Wolfe et al., 2019). Theoretically, this is true of other cues to absolute distance as well (such as accommodation and motion parallax) which is why "conventional wisdom" has traditionally identified "eye vergence, accommodation (focusing the image), binocular disparity, and motion parallax" as the four "primary cues" to depth (Rogers, 2017). See Bishop & Pettigrew (1986) for an optimistic account of 3D vision without the need to infer 3D content from the 2D retinal images, and Clark & Yuille (1990), Ch.1 for a skeptical one.

2. Computer Vision: If you were to reverse engineer distance estimates for a visual system based on two rotating cameras (or eyes), vergence would seems like the natural solution. Indeed, vergence played a central role in the 'active vision' revolution in computer vision in the late 1980s and early 1990s (see Krotkov & Kories, 1988; Krotkov, Fuma, & Summers, 1988; Abbott & Ahuja, 1988; Geiger & Yuille, 1989; Krotkov, 1989; Krotkov, Henriksen, & Kories, 1990; Abbott & Ahuja, 1990; Olson & Coombs, 1991; Blake & Yuille, 1992, esp. Ch.8: Brown et al., 1992; Coombs

& Brown, 1992; Coombs & Brown, 1993; Krotkov & Bajcsy, 1993; see also Schechner & Kiryati, 2000 for an influential discussion of distance triangulation in computer vision).

3. Effectiveness: Vergence is thought to be particularly effective compared to other absolute distance cues, such as accommodation and motion parallax. In their systematic review of the literature Thompson, Fleming, Creem-Regehr, & Stefanucci (2011) identify just four key absolute distance cues: accommodation, vergence, height in the visual scene, and familiar size. They leave a '?' next to motion parallax. (An evaluation they confirm and further justify in Creem-Regehr, Stefanucci, & Thompson, 2015). Similarly vergence is the principal absolute distance cue discussed by Vishwanath (2014; 2019) and Rogers (2019) in their recent debate on 3D vision, with Rogers (2019) asserting: "No one would deny that binocular disparities and eye vergence are sufficient to 'specify perceived depth relations'". Indeed, Rogers (2019) identifies just two cues to absolute distance: vergence and vertical disparities.

Cutting & Vishton (1995) (a standard reference in contemporary textbooks; see Goldstein & Brockmole, 2016; Thompson et al., 2011) suggest that oculomotor cues "could be extremely effective in measuring distance, yielding metric information within near distance". Empirical evidence for this claim dates back to Meyer (1842), Wheatstone (1852), and Baird (1903). Swenson (1932) found a relationship of $y = x - 0.15$ with hidden-hand pointing for distances between 25cm to 40cm, with an average error of 0.17cm, whilst Von Hofsten (1976) found a relationship of $y = 0.9x + 8.5$ for distances between 60cm and 118cm, with an average error of 2.2cm. Foley (1980) analysed a series of binocular depth distortions (depth constancy in binocular stereopsis, curvature of the fronto-parallel plane, inability to bisect distances) and argued that they all originated from the same misestimation of distance from vergence. Whilst Foley (1980) helped to cement vergence as an effective absolute distance cue, it also implied that vergence was 'non-veridical' (with the visual system's estimate of the vergence angle being only half its true value), and became the received wisdom for the next two decades.

Foley (1980) set the tone for more recent debates, starting with Mon-Williams & Tresilian (1999), where there is no question whether vergence is an effective absolute distance cue. The only question is whether vergence is veridical or not. Mon-Williams & Tresilian (1999) found a strong linear relationship between vergence and hidden-hand pointing to the distance of a point of light of $y = 0.86x + 6.5$ for distances between 20cm and 60cm (see Fig.17). In this paper, and in Mon-Williams, Tresilian, McIntosh, & Milner (2001), they therefore challenge Foley (1980)'s contention that vergence is a non-veridical distance cue. Instead, they suggest that any compression in their results (a gain of 0.86, rather than a gain of 1) is due to a cognitive strategy that subjects adopt to slightly hedge their bets towards the mean (Poulton, 1980; Poulton, 1988).

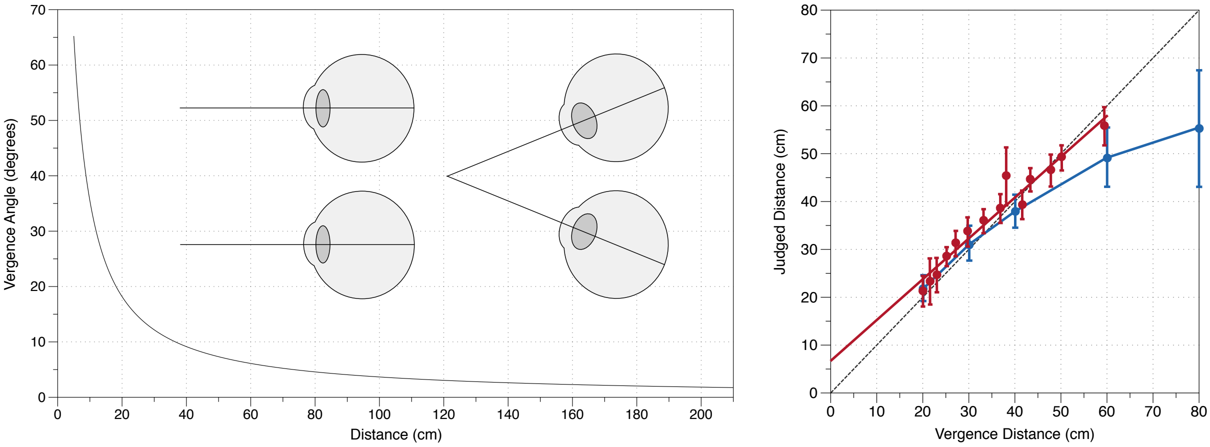

Figure 25. Distance from Vergence. Left panel illustrates how the vergence angle changes with fixation distance. Right panel illustrates the results of absolute distance from vergence in Mon-Williams & Tresilian (1999) (in red), and Viguier, Clément, & Trotter (2001) (in blue), compared to veridical performance (the black dotted line).

Viguier, Clément, & Trotter (2001) is another influential study. They presented subjects with a 0.57° disc at distances between 20cm and 80cm for 5s, and then after 5s in darkness asked subjects to match the distance with a visible reference. They found subjects were close to veridical for

20cm, 30cm, and 40cm, but distances were increasingly underestimated beyond that (60cm was judged to be 50cm, and 80cm was judged to be 56cm; see Fig.17). They conclude that:

> "…in agreement with previous studies (Von Hofsten 1976; Foley 1980; Brenner & van Damme 1998; Mon-Williams & Tresilian 1999; Tresilian et al. 1999) the results of our experiment indicate that vergence can be used to reliably evaluate target distance. This is particularly effective in the near visual space corresponding to arm's length."

Whilst Viguier et al. (2001) confirms the effectiveness of vergence as an absolute distance cue, the underestimation of distances beyond 40cm appears to challenge the suggestion that the distance extracted from vergence is veridical. In response, Scarfe & Hibbard (2017) ask whether even this underestimation can "in some senses be considered optimal?" Although Mon-Williams & Tresilian (1999) and Tresilian et al. (1999) observe that symmetric angular noise in the vergence signal will skew the range of probable distances asymmetrically towards further distances, Scarfe & Hibbard (2017) show that this implies that the most likely distance actually reduces, explaining the underestimation of distance we observe in Viguier et al. (2001). For instance, symmetric angular noise as modelled by Scarfe & Hibbard (2017) for a 60cm target is consistent with the target being between 30cm and 140cm. The distance range is skewed asymmetrically towards further distances (30-60cm vs. 60-140cm) but, and this is the key point, because this is symmetric angular noise, there is a 50% chance the target is between 30-60cm and a 50% chance the target is between 60-140cm. For this to hold true whilst the distance range is skewed towards longer distances, i.e. for there to be the same volume under the curve of the probability density function, the peak of the probability density function has to be skewed towards shorter distances (in this case 50cm).

However, there is an alternative explanation for the distance underestimation in Viguier et al. (2001). Viguier et al. (2001) use an extended stimulus with a constant angular size, which might act as a counter-cue to distance from vergence. Since a drop-off in performance is not observed after 40cm in Mon-Williams & Tresilian (1999) when a dot (with no discriminable angular size) is used, this appears to be the most likely explanation. So arguably Mon-Williams & Tresilian (1999)'s suggestion that vergence is a veridical cue at near distances still stands.

4. Peripersonal vs Extrapersonal Space: Because vergence falls off with the tangent of the distance (Fig.25), there is little change in the vergence angle beyond 2m, and it is commonly suggested that vergence's effective range doesn't extend much beyond this (Collewijn & Erkelens, 1990; Cutting & Vishton, 1995; Mon-Williams & Tresilian, 1999; Howard, 2012). One exception is Rogers (2019) who suggests: "The vergence signal indicates viewing at a large distance … signalling that the objects in the scene are (and are seen to be at) a large distance away" (see also Brenner & van Damme, 1998's suggestion that vergence can scale the distance of a bird in the sky). Christopher Tyler has also indicated in personal communications that vergence may be effective for far distances. And previously vergence was assumed to be effective at far distances, for instance Walls (1942) asserted that:

> "The amount of convergence, evaluated quite unconsciously via kinesthetic reception from the internal rectus muscles, is a potent cue to distance. It is effective up to the greatest distances for which we converge at all appreciably-up to a hundred feet or more, which is far beyond the distances for which we accommodate."[1]

One reason for believing this, is that disparity scaling is thought to be effective beyond 2m, so the distance information required for disparity scaling (vergence) must be effective beyond 2m. I.e. something is providing the absolute distance information required to extract depth from disparity, and I know of no discussions suggesting that this information comes from e.g. the ground plane.

---

[1] Thank you to Jenny Read for bringing this quotation to my attention.

Indeed, as I have already noted, the slant of the ground plane appears to be dictated by disparity itself, so we would find ourselves in a circular argument if this were the source of information.

But the importance of vergence as an absolute distance cue is especially apparent if we believe that there is an important distinction between near (peripersonal) space and far (extrapersonal) space. This segmentation of visual space was a defining feature of Cutting & Vishton (1995)'s influential review of absolute distance cues, and continues to influence the debate with the suggestion that "vergence of the eyes may provide a key signal for encoding near space" (Culham et al., 2008), and Creem-Regehr et al. (2015):

> "Binocular stereo provides accurate absolute distance information only in personal space, where it functions to support reaching. Eye-height-scaled perspective is ineffective in both personal space and vista space, but can support accurately scaled egocentric distance judgments in action space, where it helps to control locomotion."

5. Reaching & Grasping: Vergence is regarded as the preeminent absolute distance cue for reaching and grasping. Bradshaw et al. (2004) find that "vergence information dominates the control of the transport [reaching] component with minimal contribution from pictorial cues", and suggest that their results confirm Mon-Williams & Dijkerman (1999). Mon-Williams & Dijkerman (1999), Mon-Williams et al. (2001), and Melmoth, Storoni, Todd, Finlay, & Grant (2007), used prisms to manipulate vergence and demonstrate its effect on reaching. Mon-Williams et al. (2001) find that patient DF's (visual form agnosia) pointing responses almost perfectly mapped the vergence manipulation ($y = 1.00x + 2.8$ for base-in prism, $y = 0.99x + 0.22$ for no prism, and $y = 0.96x + 0.6$ for base-out prism). Culham et al. (2008) also cite earlier behavioural studies that "suggest that eye position and vergence play an important role in the accuracy of reaching movements (Bock, 1986; Henriques & Crawford, 2000; Henriques, Klier, Smith, Lowy, & Crawford, 1998; Henriques, Medendorp, Gielen, & Crawford, 2003; Neggers & Bekkering,

1999; van Donkelaar & Staub, 2000).” Other recent studies that either explore or assume the pre-eminence of vergence as an absolute distance cue for reaching and grasping include: Naceri, Chellali, & Hoinville (2011) (who found results similar to Viguier et al., 2001 for reaching and grasping a fixed-angular sized object in virtual reality); Naceri, Moscatelli, & Chellali (2015); Klinghammer, Schütz, Blohm, & Fiehler (2016); Campagnoli, Croom, & Domini (2017); Grant & Conway (2019); Campagnoli & Domini (2019).

6. Brain Imaging: Brain imaging studies also suggest that vergence acts as the primary absolute distance cue for reaching. Quinlan & Culham (2007) found that the dorsal parieto-occipital sulcus (dPOS) demonstrates a near-space preference, with high activation at closer distances. Importantly, they found that this activation arose when oculomotor cues (vergence, accommodation) were the only cues to absolute distance, leading Quinlan & Culham (2007) to conclude that “it appears that humans do have a functional area that can reflect object distance based on oculomotor cues alone.” This finding was significant for another reason, namely that the same region, the superior parieto-occipital cortex (sPOC) is “primarily – if not, exclusively – concerned with the automatic encoding of target information needed for planning the reach (Pisella et al., 2000; Gallivan et al., 2009; Lindner et al., 2010; Vesia et al., 2010; Glover et al., 2012)” (Grant & Conway, 2019; see Culham et al., 2008 for earlier literature), leading Quinlan & Culham (2007) to conclude:

> “To summarize, in the context of earlier literature, our findings suggest that near vergence is coded in dPOS, a region within the dorsal pathway that plays a critical role in reaching, particularly when the target is off-fixation. Eye position signals related to the current degree of vergence in dPOS likely supply the dorsal stream with critically important information about object distance with respect to current gaze.”

This work complements single-cell recordings that identified vergence coding in LGN (Richards, 1968), the visual cortex (Trotter, Celebrini, Stricanne, Thorpe, & Imbert, 1992; Trotter, Stricanne, Celebrini, Thorpe, & Imbert, 1993; Trotter, Celebrini, Stricanne, Thorpe, & Imbert, 1996; Masson, Busettini, & Miles, 1997; Dobbins, Jeo, Fiser, & Allman, 1998; Trotter & Celebrini, 1999), and the parietal cortex (Gnadt & Mays, 1995). These studies were inspired by the fact that the "psychophysical data suggest an important role for vergence" (Trotter et al., 1992); something that Trotter himself confirmed in Viguier et al. (2001), illustrating the important interplay between the neural and psychophysical data on this topic (see also Lehky, Pouget, & Sejnowski, 1990 for an early neural network model of vergence scaling that Trotter et al. 1992 complements).

7. Size Constancy: We cannot divorce the importance of vergence as an absolute distance cue from the central role vergence is supposed to play in scaling the size (size constancy) and 3D shape (depth constancy) of objects. On size constancy, Combe & Wexler (2010) refer to "the common notion that size constancy emerges as a result of retinal and vergence processing alone" (although they suggest that motion parallax can also have a role to play). The role of vergence in size constancy is particularly acute in the Taylor illusion (scaling an after-image of the subject's hand as the hand is moved forward and backwards in complete darkness), with Taylor (1941) and Mon-Williams, Tresilian, Plooy, Wann, & Broerse (1997) arguing that vergence is solely responsible for the illusion, and Ramsay, Carey, & Jackson (2007) and Sperandio, Kaderali, Chouinard, Frey, & Goodale (2013) only weakly qualifying that conclusion, with Ramsay et al. (2007) observing: "Of course, vergence provides an extremely powerful distance cue", whilst Sperandio et al. (2013) find that "perceived size changes mainly as a function of the vergence angle of the eyes, underscoring its importance in size-distance scaling."

8. Depth Constancy: Vergence is also thought to play a central role in the scaling of 3D shape. As Thompson et al. (2011) note, "disparity is usually considered a relative depth cue, distinct from vergence." As we have already discussed in the context of Rogers & Bradshaw (1995), when

the size of the object is less than 20° the scaling of binocular disparity is dominated by vergence rather than by vertical disparities.

Summarising the contemporary literature, then, there seems to be little question that vergence is one of our most important absolute distance cues for near distances. The consensus seems to be that "as targets get nearer, vergence information plays an increasingly important role in distance perception", with vergence providing "critically important information" in reaching and grasping (Quinlan & Culham, 2007). The only debate is whether vergence provides us with 'veridical' (or in some sense 'optimal') absolute distance information within reaching space? (Mon-Williams & Tresilian, 1999; Mon-Williams et al., 2001; Scarfe & Hibbard, 2017).

4. Experiment 1 (Vergence as a Distance Cue)

However, to the best of our knowledge vergence has never been tested as an absolute distance cue divorced from obvious confounding cues such as binocular disparity. We can therefore have little confidence that it is vergence that is determining the absolute distance in these experiments rather than these confounding cues.

We identify three confounding cues, all of which are introduced by the stimulus presentation in Mon-Williams & Tresilian (1999) and Viguier et al. (2001). Subjects are sat in complete darkness, with their vergence in a resting state, and then the stimulus is suddenly presented, often as close as 20cm. This introduces three confounding cues:

1. Double Vision (Retinal Disparity) Before Vergence: If the observer's vergence is in a resting state, and a stimulus is presented as close as 20cm, then it is going to be seen as double before the observer makes their vergence eye movement. But we know from Morrison & Whiteside (1984) that diplopia (double vision) can be an effective absolute distance cue. Morrison & Whiteside (1984) found that 90% of performance in estimating the distance of a point of light between 0.5m-9.2m could be attributed to diplopia, since performance was only degraded by 10% when the stimuli were shown for a brief period (0.1-0.2s; too quick for a vergence response). Although characterised as 'coarse stereopsis' by Ogle (1953), recent literature has emphasised how diplopia provides a direct perception of depth, rather than merely being a cognitive cue (Ziegler & Hess, 1997; Lugtigheid, Wilcox, Allison, & Howard, 2014).

One question is how diplopia might function as an absolute distance cue. The answer is that participants in darkness will know roughly where they're looking, and therefore the distance of the initial fixation plane. This isn't a claim about participants being able to monitor dark vergence, or attributions to a specific distance tendency. There's a tendency in the literature to treat initial eye position as something the participant has to monitor in order to know about, when in reality we are in control of our own eye movements. We know where we are looking, roughly speaking, because we control where we are looking look. Participants are in an experimental apparatus, stare out into the darkness, knowing that they are looking at a far distance. Any diplopia will indicate reduction in distance from this far fixation. This is not to say that the distance estimates will be well calibrated. For instance the distance judgements in Morrison & Whiteside (1984) showed only a small gain. But this explains how, with additional knowledge of the distance range of the targets (from the experimental set-up of pointing within arm's reach), participants could build up the kind of performance we observe in Mon-Williams & Tresilian (1999) over the course of an experiment (where it becomes apparent which are near, middle, and far targets).

2. Changing Retinal Image (Motion on the Retina) During Vergence: The second confounding cue is the motion of the target on the retina (as it moves from the retinal periphery to the fovea) during vergence. When the stimulus is an isolated target viewed in darkness (as it was in Mon-Williams & Tresilian, 1999 and Viguier, Clément, & Trotter, 2001), subjects will literally watch the targets in each eye streak towards each other across the visual field. Given plausible assumptions about our vergence resting state (that our vergence is beyond arms reach when our eyes are relaxed), the motion of the target across the visual field could be used to inform subjects about the absolute distance of the target.

3. Conscious Awareness of Eye Movements During Vergence: If subjects have to make a sudden vergence eye movement in a response to a near target, they will be consciously aware of their own eye movements because they will literally feel their eyes rotating. If subjects have little or no other absolute distance information, they are going to attend to these consciously felt muscular sensations and attach a lot of weight to them. But this is not how we judge distances in everyday viewing (cf. Berkeley, 1709 who argued that it is). Instead, the suggestion in the literature is that the visual system unconsciously processes muscle movements that we don't notice (sub-threshold extraocular muscle proprioception) or eye movement plans we don't know about (efference copy). Consequently, it is important to focus on sub-threshold vergence eye movements

(eye movements that subjects don't notice) if we are to get a better understanding of how vergence actually contributes to distance perception everyday viewing.

This is an important distinction which is further explored in Section 3 of the Discussion below ('3. Conscious Awareness of Eye Movements'). To give an illustration of why this is an important distinction, ever since Trotter et al. (1992) found that a large majority of neurons in the primary visual context (V1) were modulated by the viewing distance provided by vergence, it has been suggested that processing of the vergence signal in the visual cortex plays an important role in visual perception (Trotter et al., 1993): "The finding that visual responsiveness is modulated by the viewing distance as early as the primary visual cortex indicates that integration of information from both retinal and extraretinal sources can occur early in the visual processing pathway for cortical representation of three-dimensional space." The assumption is that the vergence signal is processed automatically in the primary visual cortex and other brain regions, not that this is the cortex processing our own conscious awareness of our own eye movements. So the assertion that conscious awareness of our own eye movements is not how we judge distances in everyday viewing isn't just a commitment of my account, it is also a commitment of the literature itself. As I mention, the only discussion I am aware of, of someone suggesting that conscious awareness of our own eye movements is how we judge distances in everyday viewing is Berkeley (1709).

To summarise, the extensive literature on vergence as an absolute distance cue tests vergence in the presence of an obvious confounding cue (binocular disparity), and in a way that is divorced from everyday viewing (conscious awareness of eye movements).

The concern that binocular disparity might actually explain absolute distance from vergence is not a new one. It formed the basis of Hillebrand (1894)'s critique of Wundt (1862). So it is worth pausing to ask why, a century on, this concern has never been addressed. The answer is that we seem to be faced with an intractable dilemma. Vergence eye movements are driven by diplopia. So, in order to drive a change in vergence, we have to introduce the very confounding cue that we ought to be controlling for. The solution we adopt in this article is to introduce sub-threshold changes in disparity in order to drive vergence (disparity visible to the observer's visual system), whilst keeping diplopia invisible to the observer (disparity subjectively invisible). In order to achieve this solution we have to manipulate the observer's vergence gradually, leaving us open to the objection that we are varying vergence too gradually. We address this concern in the discussion. But we highlight from the outset that this is a necessary trade-off that we have intentionally made. To test vergence as an absolute distance cue in the presence of above-threshold disparities is not an option, and we can have no confidence in experimental results that test vergence as an absolute distance cue in this way.

Experiment 1 therefore replicates the fundamental aspects of Mon-Williams & Tresilian (1999)'s experimental set-up, whilst controlling for these three confounding cues in two fundamental ways: First, to avoid any sudden jumps in vergence demand, vergence was varied between each trial by having subjects cross-fuse a fixation target, and slowly increasing or decreasing the separation between the target for the left eye and the target for the right eye over 15s. Second, the amount that the vergence demand was varied over these 15s was a modest 1.6º–1.9º (one step in a random walk over seven distances: 20cm, 22cm, 25cm, 29cm, 34cm, 40cm, 50cm), rather than up to the 17.6º of vergence demand from the sudden presentation of a stimulus 20 cm away to an observer sitting in darkness with their vergence at a resting state.

Observers

12 observers (9 male, 3 female; ages 28-36, average age 31.2) who had indicated their interest in volunteering for the study in response to a Facebook post participated in the experiment. All subjects were naïve as to the purpose of the experiment. Observers either did not need visual correction or wore contact lenses (no glasses). All observers gave their written consent, and the study was approved by the School of Health Sciences Research Ethics Committee, City, University of London in accordance with the Declaration of Helsinki.

Three preliminary tests were performed to ensure that (a) the subjects' arm reach was at least 60cm, (b) their convergence response was within normal bounds (18D or above on a Clement Clarke Intl. horizontal prism bar test), and (c) that their stereoacuity was within normal bounds (60 sec of arc or less on a TNO stereo test). 1 additional observer (LN) was excluded as he experienced diplopia from the outset of the experiment.

Apparatus

Since the purpose of the experiment was to replicate the main aspects of Mon-Williams & Tresilian (1999)'s experimental set-up, a viewing box similar to theirs (45cm high, 28cm wide, and 90cm long) was constructed: see Fig.26. The insides of the box were painted with blackout paint mixed with sand in accordance with the instructions of the optical instrument maker Gerd Neumann Jr (reference in bibliography).

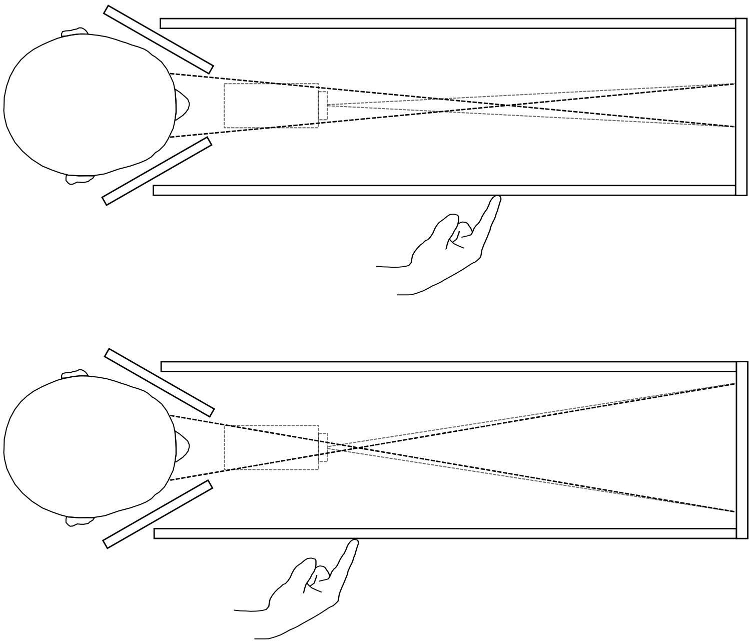

Figure 26. Apparatus for Experiment 1 viewed from above. Laser projector (in grey) projects two stimuli onto a black metal plate at the end of the apparatus. Occluders ensure that the left eye sees

only the right stimulus and the right eye sees only the left stimulus (black dotted lines). Vergence angle is varied by increasing the distance between the two stimuli (compare top and bottom images). Subjects point to the perceived distance on a ruler on the side of the box.

I altered Mon-Williams & Tresilian (1999)'s experimental set-up in two fundamental ways:

1. Stimulus alignment: Mon-Williams & Tresilian (1999) aligned their stimulus with the optical axis of the right eye, and varied the vergence demand of the left eye using prisms. This approach has two shortcomings: First, it leads to an asymmetric vergence demand. For a 20cm target aligned with the right eye, the vergence demand is 17.75° for the left eye and 0° for the right eye, rather than 8.81° for each eye had the target been located along the midline. In normal viewing conditions head rotations would eradicate such extreme asymmetries in vergence demand. Second, the stimulus is liable to be perceived as drifting rightwards as it gets closer: at 50cm a stimulus aligned with the right eye is offset from the subject's midline by 3.5°, whilst at 20cm it is offset by 8.8°. I therefore aligned the stimulus with the subject's midline.

2. Stimulus presentation: As I have already discussed, rather than varying the vergence demand with prisms, which introduce sudden jumps in vergence, I had subjects cross-fuse a fixation target and varied vergence by slowly increasing or decreasing the separation between the targets for the left and right eyes. Occlusion barriers were used to ensure each eye only saw its appropriate stimulus. The stimuli were presented using a Sony MP-CL1A laser projector, fixed 25cm in front and 5cm below the line of sight, which projected the stimuli onto a black metal plate that formed the back wall of the apparatus. Laser projection was used to ensure the stimuli were viewed in perfect darkness (unlike CRT or Liquid Crystal displays, lasers projectors emit no light for black pixels, ensuring perfect black values and no residual luminance).

Subjects were completely naïve about the experimental set-up: (a) they did not see the room or apparatus beforehand, which was sealed off by a curtain, and (b) they were wheeled into the experimental room wearing a blindfold, their hand was guided to a head and chin rest, and they had to ensure their head was in place, with a further hood of blackout fabric pulled over their head, before they could take the blindfold off. This procedure, coupled with the fact that the box was sealed, and the illumination in the room outside the box was reduced to a low-powered LED, ensured subjects viewed an isolated stimulus in perfect darkness.

Before the start of each trial subjects were asked to indicate how many targets they could see, and the metal occluders were adjusted by the experimenter until only one target was visible in each eye. The experimenter then guided their arm to a ruler, attached at eye level on the right-hand side of the box (offset to the right of the subject's midline by 12cm), and subjects were told to indicate their distance judgements on the ruler, but to relax their arm by their sides between trials. Distance judgements were recorded by the experimenter.

The stimuli, also presented at eye level, were produced in PsychoPy (Peirce, 2007; 2009). They comprise of (a) a fixation target (300 green dots located within a 1.83° circle, with the dots randomly relocating every 50ms within the circle, giving the impression of a disc of shimmering dots): see Fig.27, and (b) a single green dot that subjects had to point to. The fixation target changed in size (sinusoidally between 1.83° and 0.91° at 1Hz). The shimmering and constantly changing size of the fixation target ensured that as the vergence demand was varied, any residual motion-in-depth from retinal slip would be hard to detect.

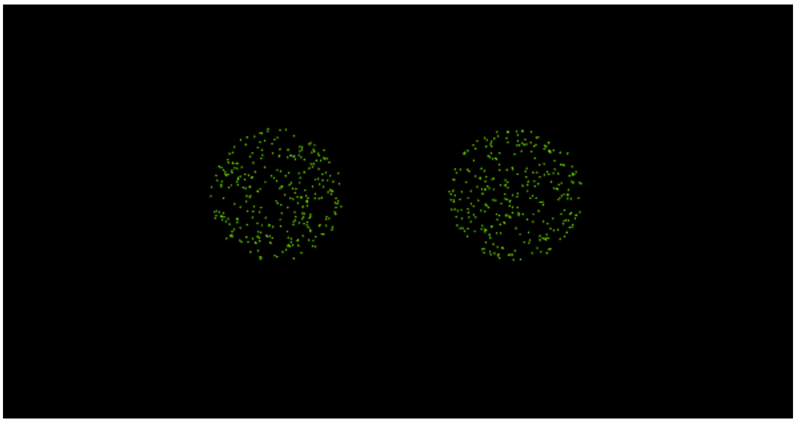

Figure 27. The fixation target in Experiment 1, comprising of 300 dots with their location within the circle updating randomly every 50ms.

Subjects completed 4 sets of 24 trials, with 96 trials in total per observer. The initial stimulus was the fixation target with the vergence angle specified for 50cm. Once subjects confirmed that the stimulus was fused, the target changed in size sinusoidally between 1.83° and 0.91° at 1Hz for 32s. Unbeknownst to the subjects, their vergence angle was also slowly increased from a vergence specified distance of 50cm to a vergence specified distance of 29cm. The stimulus then changed to a dot and subjects had to indicate its distance by pointing on the side of the box. At this point the stimulus changed back to the fixation target for 15s, and vergence was stepped up or down for the next trial using a pseudo-random walk. The pseudo-random walk covered 7 distances: 20cm, 22cm, 25cm, 29cm, 34cm, 40cm, 50cm and specified the vergence distance for the remaining 23 trials of each set. The vergence demand of the next trial was either stepped up or stepped down by one step. The walk was only pseudo-random as there was a slightly higher probability (0.6) that the step would be away from the middle of the range, to ensure full coverage of the 7 distances. Walks were simulated prior to the experiment, and a walk was chosen for each subject that ensured each of the 7 distances was tested at least 10 times over the 96 trials.

After the main experiment was completed, a control study was run in full-cue conditions to confirm Mon-Williams & Tresilian (1999)'s and Swan et al. (2015)'s findings that manual pointing with a hidden hand is a good proxy for perceived distance. The control replicated the head and chin rest, right-hand wall, and hidden ruler of the original apparatus in Experiment 1, but removed the top, back, and left-hand side of the box, enabling a familiar object (a 510g Kellogg's Rice Krispies box) to be seen in full-cue conditions. Subjects pointed to the front of the cereal box with a hidden hand in 3 sets of trials that ensured 10 trials in total for each of the 7 distances (20cm, 22cm, 25cm, 29cm, 34cm, 40cm, and 50cm). One subject (SM) was unable to return to complete the control.

Results

The results of the main experiment and the control are shown in Fig.28, with the results of the main experiment in black and the results of the control in grey.

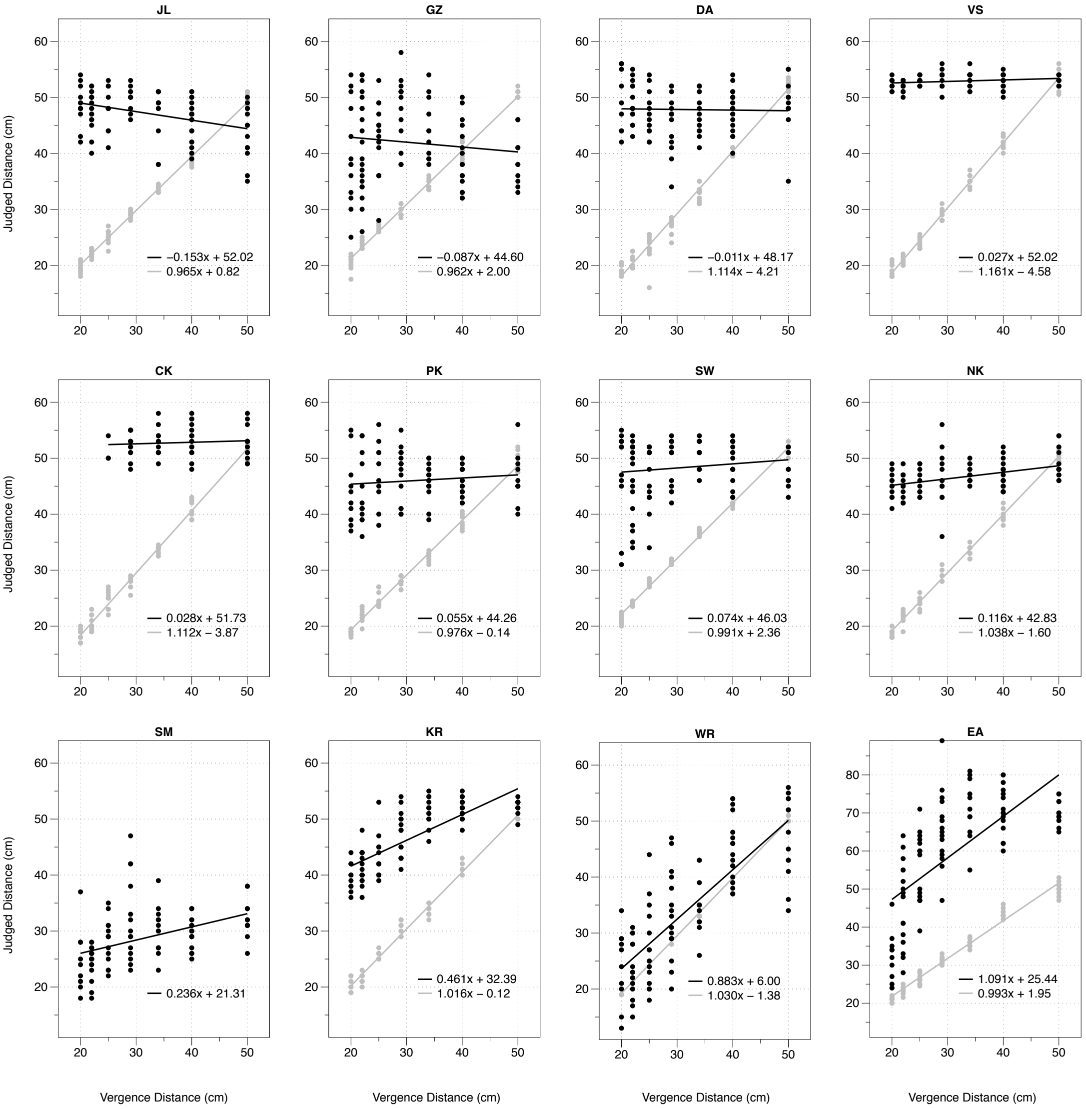

Figure 28. Results of Experiment 1. Grey dots show the results for the full-cue pointing control. Black dots show the results for the vergence-only condition.

In the control, all subjects were close to veridical when it came to pointing with a hidden hand at the distance of a familiar object in full-cue conditions. A linear mixed effects model (Pinheiro & Bates, 2000) conducted in R (R Core Team, 2012) using the lme4 package (Bates, Maechler, & Bolker, 2012; see Lisi, 2015) estimates the relationship as $y = 1.032x - 0.76$ (with 95% confidence intervals of 0.992 to 1.071 for the slope, and –2.36 to 0.73 for the intercept). This result confirms the findings of the control tests in Mon-Williams & Tresilian (1999) ($y = 1.08x - 1.35$) and Swan, Singh, & Ellis (2015) ($y = 1.005x - 2.76$) that hidden hand pointing is an accurate way of reporting perceived distance.

Turning to the results of the main experiment, we find significant individual differences in the effectiveness of vergence as a distance cue: 8 subjects with no gain whatsoever, 2 subjects with very modest gains (SM: 0.24; KR: 0.46), and 2 subjects with very significant gains (WR: 0.88; EA: 1.1, although EA consistently overshoots by 25cm). To try and make sense of these individual differences, we can cluster the histogram of the slopes in Fig.28 using a Gaussian mixture model in R with the mclust5 package (Scrucca, Fop, Murphy, & Raftery, 2017). According to the Bayesian information criterion (BIC), a model with two populations and equal variance best fits the data: one population with 10 subjects and an average slope of 0.074, and another population with 2 subjects and an average slope of 0.983: see Fig.29. We can also analyse these two populations using a linear mixed effects model: the first population of 10 subjects has a relationship of $y = 0.075x + 43.52$ (with 95% confidence intervals of –0.035 to 0.183 for the slope, and 37.12 to 49.70 for the intercept), and the second population of 2 subjects has a relationship of $y = 0.987x + 15.72$ (with 95% confidence intervals of 0.747 to 1.219 for the slope, and –3.94 to 36.13 for the intercept).

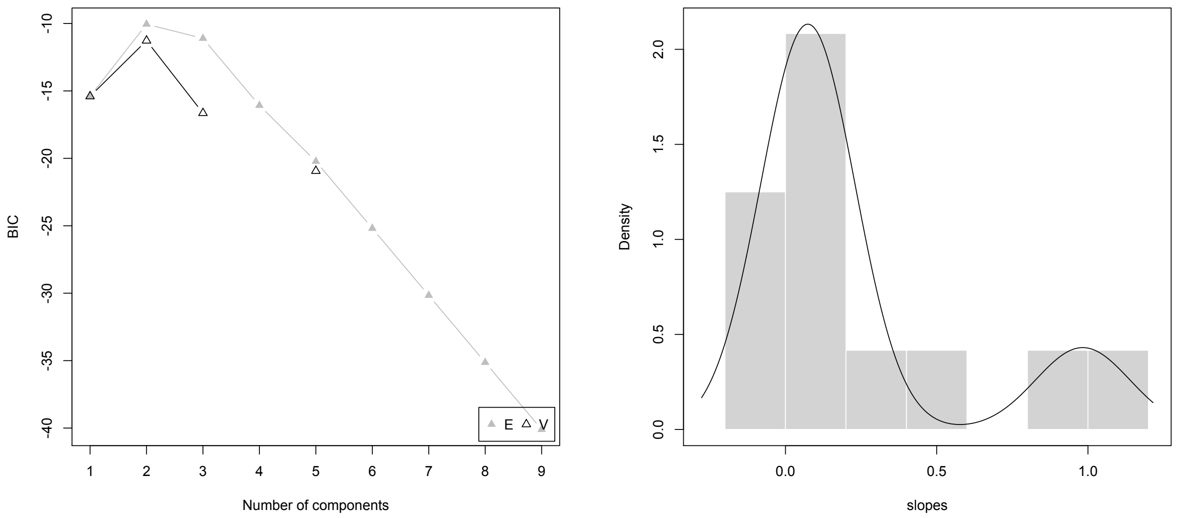

Figure 29. Clustering of the slopes in Experiment 1. Bayesian Information Criterion (BIC) indicates that two groups of equal variance best fit the histogram of the slopes in Fig.28 (left). Resulting Gaussian distributions plotted on top of the histogram of the slopes from Fig.28 (right).

Discussion

The general finding of Experiment 1 is that vergence is an ineffective distance cue for the vast majority (10 out of 12) of our subjects, with an average gain of $y = 0.075x + 43.52$ for 10 of our subjects. This suggests that the high gains reported in Mon-Williams & Tresilian (1999) ($y = 0.86x + 6.5$ between 20cm and 60cm) and Viguier et al. (2001) (close to veridical between 20cm and 40cm) can be attributed to the three confounding cues I identified in Section 2: (1) diplopia, (2) the changing retinal image, and (3) vergence as a kinesthetic cue.

But we still have to explain why the two outliers (WR and EA) have a slope close to 1? On the face of it, the stark contrast between 10 subjects with a slope close to 0, and 2 subjects with a slope close to 1, might appear to suggest that vergence is an effective distance cue for a small minority of the population. But I believe this conclusion is premature, since all three of the subjects with the highest gains (KR, WR, and EA) made comments during the experiment and/or during their post-experiment debrief that were consistent with vergence / accommodation conflict accounting for their results. The problem is that Experiment 1 induced up to 3.9D of vergence / accommodation conflict, with subjects accommodating on a back plane 90cm away, but converging to distances as close as 20cm. This conflict can manifest itself in three different ways:

First, accommodation could determine vergence, and break binocular fusion. This was only a concern for one subject (CK), who could fuse the fixation target at the closest distances (20cm, 22cm), but whose fusion broke as soon as the dot was presented. Two other subjects experience diplopia (JL twice, GZ once).

Second, vergence could determine accommodation, and the dot could go out of focus. Two of the subjects with the highest gains (KR and EA) reported this, and both reported using the change in size as the dot went out of focus as a distance cue.

Third, accurate vergence and accommodation could be maintained but at the cost, for some subjects, of significant eye strain. In a preliminary test, two very experienced psychophysical observers reported eye strain, and WR, the subject with the second highest gain, also reported eyestrain, describing the experiment as "exhausting" for his eyes. This suggests that the high gains reported by KR, WR, and EA may be explained by vergence / accommodation conflict rather than the effectiveness of vergence as a distance cue. The solution is to retest these three subjects with correct accommodation cues in place. But this poses a challenge: if vergence is being varied dynamically, how can we maintain roughly accurate accommodation cues, and test whether accommodation has any effect? This is the focus of Experiment 2.

Experiment 2 will also test 12 new participants with effective accommodation cues in place. One suggestion for the low gain of the 10 subjects in Experiment 1 is that this was due to the vergence / accommodation conflict. This seems unlikely given that vergence / accommodation conflict is a hallmark of the vast majority of studies appearing to confirm vergence as an absolute distance cue, including Mon-Williams & Tresilian (1999). But Experiment 2 tests this hypothesis.

5. Experiment 2 (Vergence and Accommodation as Distance Cues)

Experiment 2 uses trial lenses to incorporate approximately consistent accommodation cues into the experimental set-up of Experiment 1, whilst exploiting the 'zone of clear single binocular vision' (the ±1D of accommodation / vergence conflict that observers can reasonably tolerate: see Hoffman et al., 2008; Fig.30). This enables 5 vergence distances (23cm, 26cm, 30cm, 36.5cm, 45.5cm) to be tested using 3 sets of trial lenses that maintain vergence / accommodation conflict within reasonable bounds (as summarised in Fig.30):

1. Near (🔴) = –4.15D (24cm) to test 23cm, 26cm, and 30cm vergence.

2. Middle (🟡) = –3.15D (32cm) to test 23cm, 26cm, 30cm, 36.5cm, and 45.5cm vergence.

3. Far (🔵) = –2.15D (47cm) to test 30cm, 36.5cm, and 45.5cm vergence.

[Removed due to copyright]

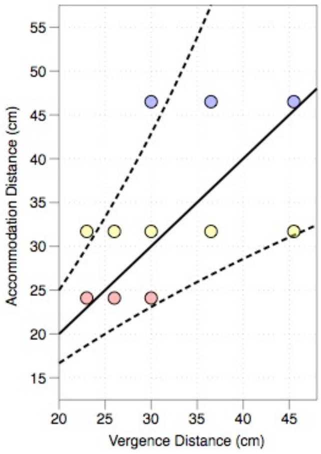

Figure 30. Vergence / Accommodation Conflict. Left: Graph from Hoffman et al. (2008) illustrating the 'zone of clear single binocular vision' (±1D). © ARVO. Right: The various combinations of vergence and accommodation tested in Experiment 2 using trial lenses. 🔴 = –

4.15D, 🟡 = –3.15D, and 🔵 = –2.15D. The solid black line indicates no vergence / accommodation conflict, and the two dotted lines mark the ‘zone of clear single binocular vision’ (±1D of vergence / accommodation conflict).

To improve the accommodative response in light of WR’s apparent difficulties, the fixation target was amended to a high-contrast random-dot 2.4° x 2.4° square, with a bulls’ eye added to the centre of the target after the first two subjects: see Fig.31 (only NM and KL used the original target; its addition is explained in the Discussion).

Figure 31. The fixation target in Experiment 2, comprising of two random-dot stereograms with a bulls’-eye at the centre of each to facilitate fixation.

As with Experiment 1, the experiment started with subjects fixating on the fixation target. Rather than changing in angular size, the fixation target varied in luminance (between 100% and 50% of its initial luminance, at 2Hz) for 50s. During this period vergence was stepped up from its initial 50cm to the middle of the specific lens’ range (26cm for Near, 30cm for Middle, and 36.5cm for Far). After the 50s, the fixation target turned into a dot and subjects had to point to its distance. As with Experiment 1, the subsequent trials continued in a pseudo-random walk, although the fixation target was visible for 30s between trials rather than 15s.

To increase contrast, the stimuli were projected onto a white screen 156cm away from the observer, rather than a black metal plate 90cm away. To ensure the increase in illumination did not also illuminate the apparatus, black fabric was added to ensure a narrow viewing window, and red filters from red-cyan stereo-glasses (blocking ≈100% green light, ≈90% blue light) were added in front of each eye. In a separate preliminary experiment using an autorefractor, the filters were found to have no impact on accommodation.

The observers were two of the observers from Experiment 1 with high gains (KR and EA) and 12 City, University of London undergraduate students (8 female, 4 male; age range 18-27, average age 20.8) recruited through flyers and Facebook Posts. All subjects were naïve as to the purpose of the experiment. The same exclusion criteria as Experiment 1 were applied, with the additional requirement that the subject's accommodative response was within normal bounds (tested using an RAF near-point rule). The study was approved by the School of Health Sciences Research Ethics Committee, City, University of London in accordance with the Declaration of Helsinki, and all subjects gave their written consent. The undergraduate students were paid £10/hr + a £20 completion bonus. 6 additional subjects had to be excluded: 4 subjects (MG, HV, FP, NS) because their prism bar convergence was below 18D, and 2 subjects (BB, RE2) because they experienced diplopia from the outset of the experiment.

The 12 undergraduates each participated in 7 sets of 20 trials (2 Near, 3 Middle, 2 Far) in random order, ensuring each lens / vergence combination was tested at least 10 times. KR and EA participated in a reduced version of this experiment with 4 sets of 20 trials (1 Near, 2 Middle, 1 Far), with only 23cm and 26cm tested in the Near condition, 26cm, 30cm, 36.5cm tested in the Middle condition, and 36.5cm and 45.5cm tested in the Far condition.

Results

The results of the two prior subjects (KR and EA) are plotted in Fig.32 against their previous performance in Experiment 1.

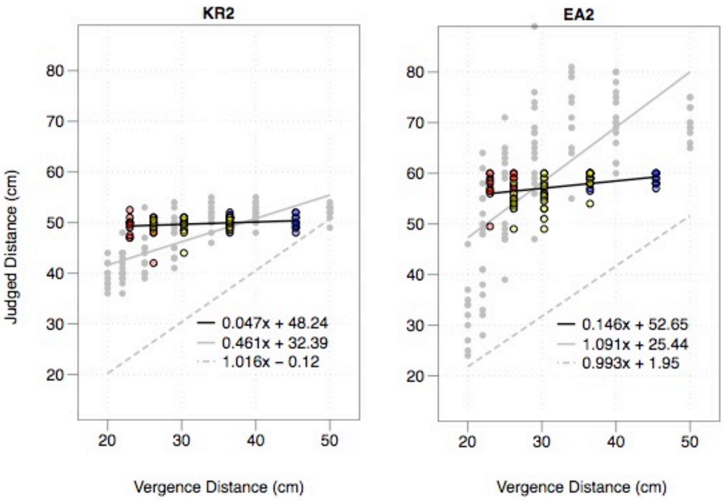

Figure 32. Performance of subjects KR and EA from Experiment 1 compared with their performance in Experiment 2. Performance in Experiment 1 indicated by grey dots and grey line. Performance in Experiment 2 indicated by coloured dots (● = –4.15D, ● = –3.15D, and ● = –2.15D) and black line. Performance in full-cue control condition indicated by broken grey line.

What is striking is how their previously high gains are eradicated with the addition of accommodation cues: KR's previous performance of $y = 0.461x + 32.39$ drops to $y = 0.047x + 48.24$, whilst EA's previous performance of $y = 1.091x + 25.44$ drops to $y = 0.146x + 52.65$ (both of these drops in performance are significant to $p < 0.001$). This confirms the hypothesis that the high gains in Experiment 1 were driven by vergence / accommodation conflict, since eliminating this conflict effectively eliminates their capacity to estimate distance. It also confirms the almost complete ineffectiveness of vergence and accommodation as distance cues for these two observers.

The results for the 12 undergraduates are plotted in Fig.33.

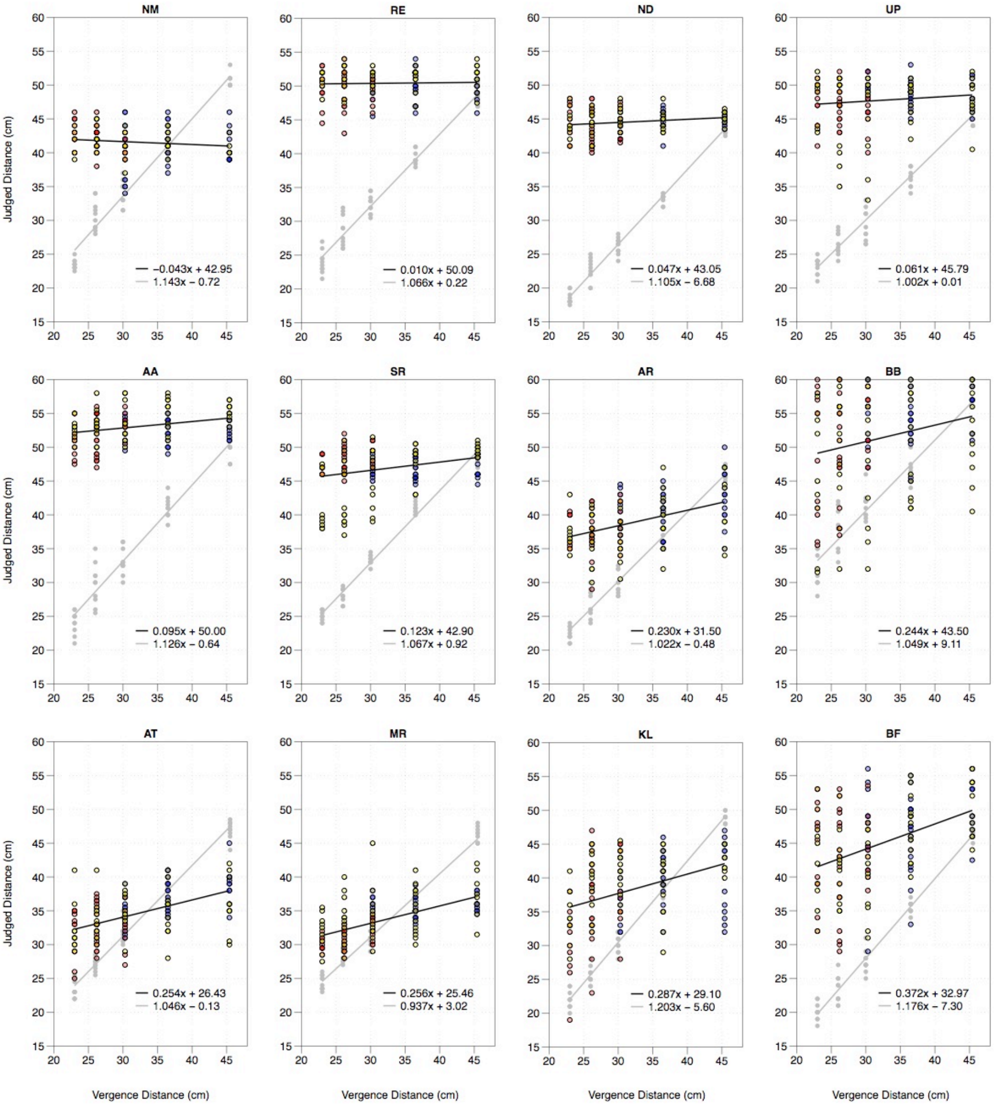

Figure 33. Performance of the 12 new subjects in Experiment 2 indicated by coloured dots and black line. Colour of dots correspond to accommodative demand (● = –4.15D, ● = –3.15D, and ● = –2.15D). Grey dots and grey line indicate performance in the full-cue control condition.

Turning to the full-cue control task, we again find a strong linear relationship $y = 1.078x - 0.69$ using a linear mixed effects model across the 12 observers (95% confidence intervals of 1.036 to 1.122 for the slope, and –3.19 to 1.81 for the intercept), further confirming hidden hand pointing as an effective reporting mechanism for perceived distance.

But turning to the main experiment, we again find significant individual differences: 6 subjects with virtually no gain, and 6 subjects with a gain of between 0.23 and 0.37. To try and make sense of these individual differences, we can cluster the histogram of the slopes in Fig.33 using a Gaussian mixture model. According to the Bayesian information criterion (BIC) we find a single population with an average gain of 0.161 best fits the data: see Fig.34, although this is only marginally better than two populations with equal variance. Fitting the data to a single population with a linear mixed effects model we find a relationship of $y = 0.161x + 38.64$ (with 95% confidence intervals of 0.090 to 0.239 for the slope, and 33.43 to 43.36 for the intercept). To put this value in context, a combined gain of 0.16 from vergence and accommodation is less than 60% of the 0.27 gain that is commonly attributed to accommodation alone as a distance cue (Fisher & Ciuffreda, 1988). Since accommodation is considered an ineffective distance cue, the same must now be concluded of vergence and accommodation in combination.

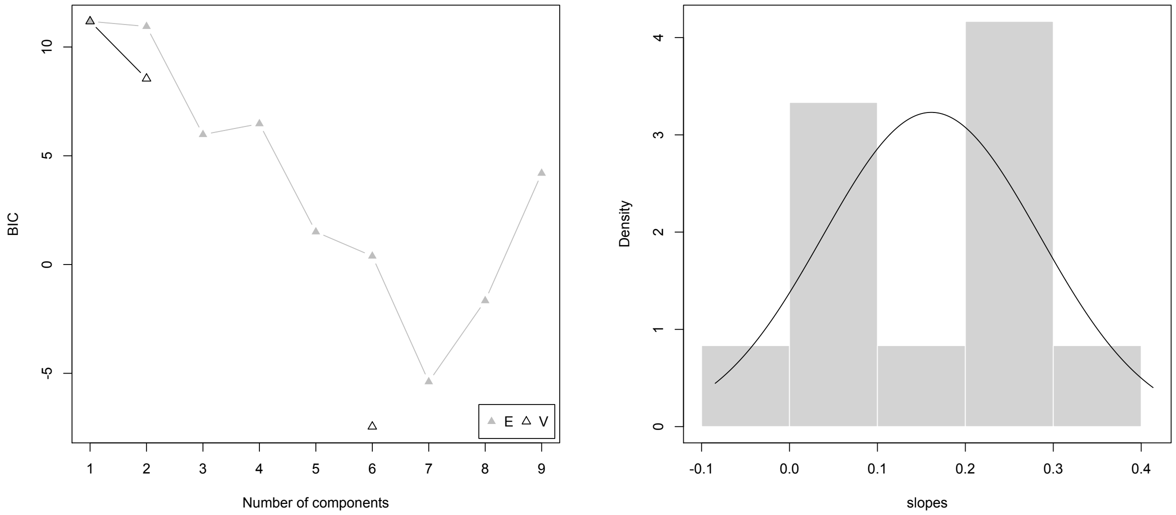

Figure 34. The Bayesian Information Criterion (BIC) indicates that a single population best fits the histogram of the slopes in Fig.33 (left), with the resulting Gaussian distribution plotted on top of the histogram of the slopes from Fig.33 (right).

This conclusion is further supported by the variance of the 6 subjects with the highest gains. We can estimate the standard deviation of the residual error (i.e. how much each subject departs from their own line of best fit in Fig.33), after correcting for motor error (assuming that perceptual error and motor error are independent) using the following formula:

$$\sigma_{residual} = \sqrt{{\sigma_{vergence}}^2 - {\sigma_{control}}^2}$$

Even if we limit ourselves to ± 2 standard deviations from the slope of best fit to rule out any outliers, we still find residual errors of 27cm for BB, 23cm for BF, 20cm for KL, 12cm for AR, 11cm for AT, and 9cm for MR, with an average of 17cm. Given the range of the experiment itself was only 22.5cm, and pointing was confined to reaching space, one is left questioning just how functionally useful cues with this degree variance could be. Mon-Williams & Tresilian (2000) come to the same conclusion in the context of accommodation alone; although the variance was higher in that study, so too were the average gains.

Finally, Experiment 2 could have been conducted with just the 'middle' lens (–3.15D), since this lens covers the full range of vergence distances (see Fig.30). Did varying accommodation by 2 dioptres using the 'near' and 'far' lens have any effect? And would the results have been any different if we eradicated the ±1D of vergence / accommodation conflict completely? We can test this hypothesis by comparing the results for the 23cm, 30cm, and 45.5cm vergence distances using (a) the 'middle' lens for all three of these vergence distances, against (b) the 'near' lens for 23cm, the 'middle lens' for 30cm, and the 'far' lens for 45.5cm, where there is no vergence / accommodation conflict. We only find a marginal reduction in performance from (a) $y = 0.176x + 38.02$ when only the 'middle' lens is used, to (b) $y = 0.147x + 38.91$ when the vergence / accommodation conflict is eradicated, and this reduction in performance is not statistically significant.

This result has two implications: First, the performance in Experiment 2 would have been no better if the vergence / accommodation conflict had been completely eradicated, indeed it might have been marginally worse. Second, the modest gain of 0.16 cannot simply be attributed to the residual vergence / accommodation conflict in the apparatus.

Discussion

These results show that vergence and accommodation were ineffective absolute distance cues for our participants. The distance estimates of some of the participants in Experiment 2 were biased by vergence. But this is not evidence of unconscious processing of the vergence signal. Instead, the subjects with the highest gains reported responding to consciously felt muscular sensations from intense sustained near fixation. For instance BF, the subject with the highest gain, reported that the experiment was "messing up my accommodation":

> "I could feel my eye are working, my eyes are focusing then relaxing then focusing."
>
> "I really had to focus to stop them going two … the target started to separate when I didn't really focus on it."
>
> "I usually get the same sensation when I'm up too late and doing some studies – a slight strain in the eye, it's not too bad, it's just that you really have to focus."

Similarly KL, the subject with the second highest gain, reported that with near targets she felt her "eyes accommodating a lot to get them to work." So, whilst our experimental paradigm effectively controlled for the conscious muscular sensations that accompany eye movements (kinaesthesia), it failed to control for the conscious muscular sensation of sustained near fixation (proprioception). These are two very distinct sensations. For instance, consider lifting a box with your arms. You have the kinaesthetic sensation of lifting the box, but after a short while, the proprioceptive sensation of the tension your muscles are under just holding the box still. The same is true with vergence / accommodation at very near fixation distances. Such muscular sensations are rarely felt in everyday viewing, and appear to be a shortcoming of manipulating vergence and accommodation as pure optical reflexes (see Charman & Heron, 2015 for similar concerns about Badal systems).

We believe that these results support the conclusion that vergence and accommodation are ineffective absolute distance cues. Specifically, that the visual system does not use vergence and accommodation to triangulate the absolute distance of objects. However, there are a number of alternative interpretations of these results that we cannot conclusively reject. In the remainder of this Discussion we therefore explore five alternative interpretations of these results that have been put to us. We explain why we do not find these alternative explanations convincing, but we recognise that none of these possibilities can be definitively excluded.

1. No Eye-Tracking

We did not use eye-tracking to track the subjects' vergence, so how do we know that subjects were actually changing their vergence during the experiment?

We did not use eye-tracking for three reasons: First, Hooge, Hessels, & Nyström (2019) found that readily available research eye-trackers "are not accurate enough to be used to determine vergence, distance to the binocular fixation point and fixation disparity", with errors of up to 2.5°. Second, we were concerned that eye-tracking would be impractical given our use of parallax barriers, making a clear view of both eyes (for the eye-tracker) and the calibration targets (for the observer) impossible. Third, we were very careful about the prior knowledge that subjects had about the apparatus, and we feared that calibration would be impossible without compromising this in some way.

Collewijn & Erkelens (1990) are critical of studies that do not provide an objective measure of vergence using eye-tracking. However, we rely on a subjective measure of vergence (diplopia), which we would argue is more reliable than camera-based eye-tracking. Before each set of trials, we asked subjects to confirm they could see the target in each eye monocularly, and then confirm that they could see a single fused target when they opened both eyes. We asked them to report if the target went double at any time during each set of 24 trials (Experiment 1) or 20 trials (Experiment 2). We paused and restarted the experiment after a break if it did. In the break between sets of trials, we also asked the subjects to confirm the target had been fused in the previous set of trials.

Since our target was a single dot, the presence or absence of diplopia provides us with a very effective test of binocular fusion. Schor & Tyler (1981) estimate diplopia thresholds for a fixation dot to be 8 arcmin. Diplopia thresholds for thin vertical bars have been found to be as low as 3 arcmin (Schor & Tyler, 1981) and 5 arcmin (Schor, Wood, & Ogawa, 1984). This helps to explain why nonius lines have traditionally been treated as a gold-standard for vergence "even when", as Schor, Wood, & Ogawa (1984) note, "small discrepancies between subjective and objective measures of horizontal fixation disparity are taken into account (Kertesz et al., 1983)." Kertesz, Hampton, & Sabrin (1983) found an average diplopia threshold of 6 arcmin for nonius lines, whilst Jaschinski, Bröde, & Griefahn (1999) found diplopia thresholds of 5 arcmin or less when measured binocularly, and 2 arcmin or less when measured with dichoptic nonius lines. In recent work Grove, Finlayson, & Ono (2014) found higher diplopia thresholds (around 13 arcmin for uncrossed disparities and 8 arcmin for crossed disparities), but their vertical bars were 4.4 arcmin wide (vs. 1.5 arcmin dot and 1.5 arcmin lines used by Schor & Tyler, 1981), so their thresholds should arguably be reduced by 3 arcmin to 5-10 arcmin for a dot stimulus.

In conclusion, a best estimate of the accuracy of diplopia thresholds in our experiment should be no more than about 10 arcmin (8 arcmin Schor & Tyler, 1981), and could well be lower if the thin vertical bar / nonius line literature applies. Compare this to objective measures from readily available research eye-trackers, where the 2D gaze literature (Choe, Blake, & Lee, 2016; Drewes, Zhu, Hu, & Hu, 2014; Wildenmann & Schaeffel, 2013; Wyatt, 2010) and the 3D gaze literature (Hooge et al., 2019) report similar errors of magnitude (up to ≈ 2.5°). Since our subjective measures are an order of magnitude (up to 15 times) more accurate than readily available objective measures from eye-tracking (10 arcmin vs 2.5º), we conclude that our subjective test for fusion based on diplopia is to be preferred. We recognise that some authors feel especially strongly that vergence studies should be accompanied by eye-tracking (Collewijn & Erkelens, 1990), but there is reasonable disagreement on this point, and notable studies share our accuracy (Hooge et al., 2019) and logistical (Quinlan & Culham, 2007) concerns.

2. Vergence-Accommodation Conflict

Could vergence-accommodation conflict account for our results? We do not believe so for three reasons.

First, as we discussed in the Introduction, there is widespread skepticism that accommodation functions as an effective absolute distance cue. Recall Mon-Williams & Tresilian (2000)'s finding that accommodation provides "no functionally useful metric distance information".

Second, vergence-accommodation conflict is a facet of most of the studies that demonstrate (close to) veridical absolute distance from vergence. First, any study which varies vergence using prisms, such as Mon-Williams & Tresilian (1999), keeps accommodation fixed. Since Mon-Williams & Tresilian (1999) varied vergence over 3.33D (20-60cm), they induce at least 1.67D of vergence-accommodation conflict, and potentially even more (we were unable to determine the exact figure). Second, any study that relies on a fixed display such as Von Hofsten (1976), is going to induce significant vergence-accommodation conflict. Von Hofsten (1976) found an almost perfect relationship between vergence and perceived distance up to 118cm, at which point there was 1.3D of vergence-accommodation conflict. Third, any study that relies on virtual reality, such as Naceri, Chellali, & Hoinville (2011), is going to induce 4D of vergence-accommodation conflict (their nearest target was 25cm, with accommodation set close to optical infinity), and yet they found results consistent with Viguier et al. (2001). So vergence-accommodation conflict hasn't previously been an impediment to finding that vergence is an effective absolute distance cue, and the maximum vergence-accommodation conflict within our second experiment (1.17D) is well within the range of these previous experiments.

Third, we explicitly tested the effect of vergence-accommodation conflict in our second experiment by contrasting the results for three vergence distances (23cm, 30cm, and 45.5cm) when (a) there was virtually no vergence-accommodation conflict (accommodation set at: 24cm, 31.5cm, and 46.5cm respectively) vs. (b) when there was up to 1.17D of vergence-accommodation conflict (accommodation set at 31.5cm) and found a non-statistically significant reduction in performance in the no vergence-accommodation conflict condition. Coupled with the fact that subjects KR and EA reduced in performance between Experiment 1 and Experiment 2 (when accommodation cues were provided), and we can conclude that vergence-accommodation conflict is not the explanation.

A related question is what drives vergence and accommodation? And does this indicate any processing of absolute distance information? First, a key thing to recognise is that in everyday viewing, a lot of eye movements are intentional. We purposefully look at a certain position (distance and direction) in space. So there is little question of how the visual system (accommodation and vergence) plans those eye movements, because the plan is internally generated in the first place. Second, when our attention is drawn to an object in full cue conditions, the distance of the object will be provided by all the distance cues available, so this becomes a restatement of the question of how we judge distances in full cue conditions. Third, the key question is therefore what determines vergence and accommodation in reduced cue conditions, for instance a single point of light in darkness. In a series of recent papers, Cholewiak et al. (2017; 2018) demonstrate how important chromatic aberration is for driving accommodation, with the finding that the accommodative response works to reduce chromatic aberration in the retinal image even at the cost of increasing blur in the retinal image. Vergence is driven by diplopia, and

Lugtigheid, Wilcox, Allison, & Howard (2014) demonstrate that observers are able to reliably judge which direction diplopia is in. So both accommodation (chromatic aberration) and vergence (diplopia) are driven by a signal that has both a direction and a magnitude component, but both of these components relate to changes off the focal plane, rather than any absolute distance cue.

3. Conscious Awareness of Eye Movements

We controlled for conscious awareness of vergence eye movements. One objection is that this is what is meant in the literature by vergence functioning as an effective absolute distance cue. We disagree with this suggestion for five reasons:

First, conscious awareness of our own eye movements is not how we judge distances in everyday viewing. Although this is an assertion, it appears to be shared by all sides of the literature. I know of no discussion apart from Berkeley (1709) that suggested otherwise. Instead, the assumption has been that absolute distance is unconsciously processed by the visual system. Subjective awareness of our own eye movements isn't what Rogers (2019) means when he suggests: "No one would deny that binocular disparities and eye vergence are sufficient to 'specify perceived depth relations'", what Cutting & Vishton (1995) mean when they suggest vergence "could be extremely effective in measuring distance, yielding metric information within near distance", what Culham et al. (2008) meant when they suggest that "vergence of the eyes may provide a key signal for encoding near space", or what Bradshaw et al. (2004) meant when they suggest that "vergence information dominates the control of the transport [reaching] component with minimal contribution from pictorial cues" in reaching and grasping tasks. These are all claims about vergence being a highly effective absolute distance cue in everyday conditions, where we don't consciously attend to our own eye movements. If vergence is only effective when subjects are consciously attending to their own eye movements, then this literature must be wrong.

Second, conscious awareness of eye movements might contribute to performance in controlled experimental conditions, which is why we controlled for them. But there is no suggestion in the literature that, even in experimental conditions, conscious awareness of eye movements could provide us with the kind of (close to) veridical estimates of absolute distance found in the experimental literature. Could subjects really achieve a relationship of $y = 0.86x + 6.5$ (Mon-Williams & Tresilian, 1999) from conscious awareness of eye movements alone?

Third, we know that the visual pathway has access to the vergence signal in LGN, the visual cortex, and the parietal cortex. There is no suggestion that what is being observed in these brain imaging studies is our conscious awareness of our own eye movements. Instead, the suggestion is that the visual system is unconsciously processing the vergence signal, and the question is whether this is actually used to provide distance information. Quinlan & Culham (2007) are not talking about conscious awareness of eye movements when they conclude: "Eye position signals related to the current degree of vergence in dPOS likely supply the dorsal stream with critically important information about object distance with respect to current gaze."

Fourth, no explanation has been given as to how conscious awareness of eye movements could explain how vergence is supposed to change our visual experience of size ('size constancy') or 3D shape ('depth constancy'). In particular, it becomes very difficult to understand how conscious awareness of eye movements could be involved in disparity scaling. Furthermore, as Regan, Erkelens, & Collewijn (1986) document, changes in size and 3D shape from vergence occur even when there is no appreciable motion-in-depth from vergence (by using large-field stimuli to veto vergence as a motion-in-depth cue).

Fifth, if subjects are merely responding to a conscious awareness of eye movements, then it is important to recognise this is not visual processing. With the potential exception of blindsight, visual processing affects our visual experience. By contrast, this account of vergence as an absolute distance cue is a purely somatosensory account that has no effect on our visual experience. To give a crude analogy, if I poke you in the eye with a pencil, you may now have a veridical sense of the absolute distance of the pencil, but it would be an aberration of language to call this visual processing. Our point is that under this alternative account, vergence as an absolute distance cue is analogous to the poking in the eye case, rather than the visual processing of absolute distance it was supposed to represent. Elsewhere we have developed this somatosensory account of vergence to encompass not just vergence as an absolute distance cue, but also vergence as a cue to motion-in-depth (Linton, 2018). The key point being that under this account, vergence does not change what we see. We look forward to developing this account to explore, and potentially encompass, directional (version) eye movements as well as depth (vergence) eye movements.

As noted, this account is in contrast Berkeley (1709). Berkeley believed we judged distances through an association between our conscious experience of our eye movements and the distance they specified. But the important thing to recognise is that Berkeley was forced into this position through his metaphysics. He thought that all there were, were "ideas". So the possibility of there being unconscious processing of any form was an anathema to Berkeley. For where would such unconscious processing occur? In the brain? But for Berkeley there is no brain in that sense, only a collection of conscious experiences and thoughts that are related to one another by associative learning. So Berkeley's argument that when I consciously experience my eye movements, I think of distance, isn't driven by introspection, let alone experimentation, but metaphysical necessity.

But let me give another analogy to try and explain what I mean when I say that vergence is not a cue that changes our visual experience. Consider how *visual* (optic flow) and *kinesthetic* (vestibular) cues interact in the experience of illusory self-motion. Vestibular cues are not necessary for illusory self-motion (Fischer & Kornmüller, 1930); for instance, when we see a neighbouring train pull away in a railway station whilst we are stationary. Nonetheless, the addition of vestibular acceleration can contribute to the illusion of self-motion (Ash et al., 2011). But, and this is the crucial point, there is no suggestion that this contribution has to be *visual*. It is not as if the optic flow has to be seen as flowing faster in order to incorporate this vestibular cue into our determination of self-motion. Instead, the visual (optic flow) and the non-visual (vestibular) cues merely feed into a common post-perceptual determination. Indeed, as Ash, Palmisano, & Kim (2011) observe, there is no 'mandatory fusion' between the *visual* (optic flow) and *kinesthetic* (vestibular) cues, which is often taken to be the litmus test of a truly perceptual effect: see Hillis, Ernst, Banks, & Landy (2002).

In the same way, I argue that our visual experience of a point of light is the same whether that point is located at 20cm or 60cm. What changes in an experimental setting is that for a near target (20cm), the initial diplopia we see, the streaking of the diplopia across our visual field as we fuse the target, and the muscular sensations we feel as we suddenly converge, inform our judgement of where the point of light is.

4. Change-Blindness

Another suggestion is that the changes in vergence in our experiments were too gradual for the visual system to detect. Similarly, in a series of 'expanding room' experiments by Glennerster and colleagues, subjects failed to notice gradual changes in vergence and motion parallax: "Subjects seem to ignore information both about vergence angle (to overrule stereopsis) and about stride length (to overrule depth from motion parallax)." (Glennerster, Tcheang, Gilson, Fitzgibbon, & Parker, 2006; see also Rauschecker, Solomon, & Glennerster, 2006; Svarverud, Gilson, & Glennerster, 2010; Svarverud, Gilson, & Glennerster, 2012). But there are two responses to this concern:

First, the failure to notice changes in vergence in the 'expanding room' experiments may have little to do with the gradual nature of the vergence change for four reasons: First, the vergence range in those experiments was limited (75cm to 3m). But we know vergence is supposed to be most effective as a distance cue within arm's reach, and Glennerster et al. (2006) merely interpret their results as indicating that the "efficacy of motion and disparity cues is greater at near viewing distances." By contrast our experiments test distances within arm's reach. Second, the 'expanding room' experiments use full-field stimuli which we know vetoes motion-in-depth from vergence, even when the change in vergence is far from gradual (up to 13.5°/s in Erkelens & Collewijn, 1985a; 1985b). Third, the pictorial cues in the 'expanding room' experiments provide the illusion of a stable scene. So all this demonstrates (as the title of Glennerster et al., 2006 illustrates) is that "humans ignore motion and stereo cues in favor of a fictional stable world". Finally, Rogers (2011) is highly critical of the 'expanding room' experiments being used as evidence of subjects failing to notice gradual vergence changes, and found conflicting results when he tested gradual vergence changes: "the gradualness of the change in interocular separation (and hence vergence demand) did not preclude the appropriate scaling of the disparity-specified ridge surfaces."

Second, even if the gradual nature of the change is responsible, subjects in the 'expanding room' experiments actually notice the change once they have been alerted to its possibility. In this regard the 'expanding room' experiments are no different from gradual colour change-blindness experiments where a region of a painting gradually changes in colour without subjects noticing (Simons, Franconeri, & Reimer, 2000; Auvray & O'Regan, 2003). As I've already explained in Linton (2017), p.102, this change-blindness is better thought of as cognitive rather than perceptual (can it really be maintained that as a region of the painting changes from red to blue over 30 seconds, the observer's visual experience remains red over the course of the 30 seconds?). But the important point is this. If you ask subjects in the gradual colour case to discriminate the colour at t_{30} (e.g. by asking them 'what colour is this region of the picture?'), they can do so accurately even though they don't detect the change during the experiment. Interestingly, what Glennerster et al. (2006)'s experiment shows is that subjects are actually very good at detecting gradual changes in vergence and motion parallax once they have been alerted to their possibility; i.e. when they re-evaluate the distances in the scene rather than simply assuming the previous depth relations in the scene apply. So in both the gradual colour case, and the gradual vergence and motion parallax case, there is an absolute signal at t_{30} that subjects have access to even if they miss the gradual change from t_1, t_2, … t_{30}. In conclusion, it would be no criticism of a colour discrimination task that the colours were gradually varied between trials. If the colour was blue at t_1, and red at t_{30}, subjects would still be able to recognise the colour at t_{30} when asked 'what colour is this region of the picture?', even though they failed to detect the colour change.

5. Delta Theta rather than Theta

One suggestion that has been put to us, is that there is a disanalogy between gradual changes in colour, and gradual changes in vergence. This argument suggests that whilst gradual changes in colour have two components (the incremental change in shade from t_1, t_2, … t_{30}, and the absolute colour at t_{30}), with subjects reporting the absolute colour at t_{30}, in the case of distance from vergence there is no absolute value at t_{30}, only the incremental changes from t_1, t_2, … t_{30}. Put another way, our experimental results have been interpreted as supporting an intermediate position. We have proved (a) that the visual system is unable to extract absolute distance from static vergence (vergence angle theta), but (b) the visual system may still be able to extract absolute distance from changes in vergence (delta theta), and the reason we don't detect this ability to extract absolute distance from delta theta is that our vergence changes are too gradual.

The claim of the delta theta account is that small changes in vergence are unconsciously integrated over time to provide us with a measure of absolute vergence. There are five responses to this suggestion:

First, it is a departure from the orthodox interpretation of vergence as an absolute distance cue. See Howard (2008), citing Swenson (1932), Mon-Williams & Tresilian (1999), and Viguier, Clément, & Trotter (2001), that: "Several studies have revealed that … people can judge the absolute distance of a visual target when the only information is provided by static vergence." Indeed, traditionally the puzzling facet of the literature was that static vergence was such an effective absolute distance cue, but by comparison motion-in-depth from delta theta was not: "the distance of a stationary object can be judged on the basis of vergence alone. So why was motion-in-depth not produced by changing vergence?" (Howard, 2012; motion-in-depth from vergence discussed in Ch.6, below). So it would now be surprising that for delta theta to be proposed as the effective absolute distance cue whilst static vergence is not.

Second, why think that the vergence changes in our experiments were too gradual for the visual system to detect? First, the visual system is clearly responding to the change in binocular disparity. As we document above, we ensured that participants did not experience diplopia, and used this as a criterion for binocular fusion. So the visual system is both (a) making vergence changes in response to a gradual target, but (b), according to this argument, nonetheless unable to monitor these very same changes to provide absolute distance. Second, these gradual changes are clearly detectable as version eye movements (horizontal motion in x-axis). This is true as monocular version signals (demonstrating Tyler, 1971's observation that "two eyes less are less sensitive than one"). But all binocular vergence is, is these two monocular version signals. Similarly, they would be detectable as a binocular version signal if they were both in the same direction. But the only difference between a binocular vergence signal and a binocular version signal is the direction of one of the signals. As Erkelens & Collewijn (1985a) note, this suggests that the visual system must have access to these binocular eye movements in order to supress them when they are equal and opposite (vergence eye movements), but not when they are equal and in the same direction (version eye movements).

Third, we have real difficulty making sense of the proposal that the visual system is able to extract absolute distance (theta) purely from a change in vergence (delta theta). As Brenner & van Damme (1998) observe, simply knowing how much the vergence angle has changed "can be of little use for judging distances if we do not know the orientation of the eyes before the change (1 deg of ocular convergence could be due to a shift in gaze from 20 to just over 21 cm or from 2 to approx. 4 m)." Clearly advocates of this position mean to suggest something more than the idea that vergence is a relative depth cue that can be scaled by an independent source of absolute distance information, otherwise every relative depth cue becomes an absolute distance cue by definition. But we really struggle to make sense of what the positive alternative is.

One suggestion is that vergence is combined with Gogel (1969)'s 'specific distance tendency', the suggestion that subjects default to a prior of 2-3m. In theory this could relate to 'dark vergence' and/or 'dark accommodation' (the natural resting state of the eyes), although (a) 'dark vergence' and 'dark accommodation' tend to be closer (around 1m for vergence, and 76cm for accommodation: see Owens & Liebowitz, 1980; Jaschinski et al., 2007), and (b) it is unclear why the visual system should have access to static vergence in the one particular context of 'dark vergence', but not more generally.

However, we should note that this suggestion is a distortion of the traditional relationship between vergence and the 'specific distance tendency' posited in the literature (Mon-Williams & Tresilian, 1999). There the suggestion is that vergence is an independent source of absolute distance information whose measurement of absolute distance is tempered by the 'specific distance tendency'. By contrast, here the suggestion is that 'specific distance tendency' usurps vergence as the independent source of our absolute distance information.

In any case there is no evidence for a tendency towards 2-3m (or 1m, or 76cm) in the vergence distance literature (see Fig.25). Mon-Williams & Tresilian (1999) find a slight contraction of the results around 40cm, not the far distances this account would predict. Viguier, Clément, & Trotter (2001) find an underestimation of distances beyond 40cm, not the overestimate that this account suggests. And given (according to this account) the apparent absence of the delta theta in our experiments, we should expect our results to be dominated by the 'specific distance tendency'. But this doesn't happen. As Fig.35 demonstrates, the key finding of our results is pervasive variance. There is no sense in which our results cluster around any specific distance.

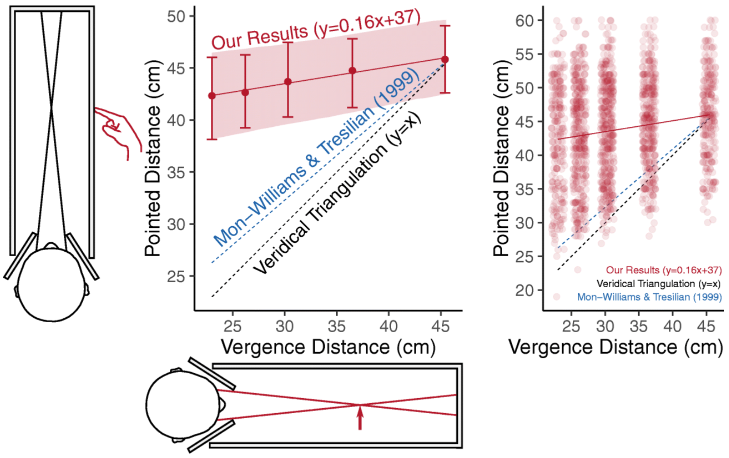

Figure 35. Summary of results in Experiment 2 (Linton, 2020). The left panel illustrates averaged results. The error bars represent bootstrapped 95% confidence intervals across observers. The error band represents the bootstrapped 95% confidence interval of the linear mixed-effects model. The right panel plots the raw trial data across observers as a jitter plot.

Intriguingly, the same pervasive variance also holds true of Gogel (1969)'s own results, where the standard deviation of distance estimates is the same size as the actual distance estimates themselves. In the presence of such pervasive variance in ours and Gogel (1969)'s results, we conclude that there is no sense in which a 'specific distance tendency' meaningfully exists.

Finally, even if the specific distance tendency were to provide the absolute distance for the initial vergence eye movement, each subsequent vergence eye movement would then have to trace its absolute distance back to this initial estimate. We'd be stuck in a near infinite regress trying to integrate over successive eye movements. If the argument is that vergence is an important absolute distance cue in everyday viewing, as opposed to single-shot distance estimates in controlled experimental conditions (such as Mon-Williams & Tresilian, 1999 and Viguier, Clément, & Trotter, 2001), then this is another serious challenge to the account.

Fourth, as one Reviewer noted, another criticism of this account is that vergence would be de-calibrated by any slow changes in everyday viewing. One response is that vergence could be recalibrated by other absolute distance cues. But it's important to recognise what this recalibration would entail. Under the delta theta account, the visual system doesn't know that vergence has gradually changed, so if vergence changes gradually from 20cm to the horizon, the recalibration would have to involve equating the old vergence angle (20cm) with the new viewing distance (the horizon). Presupposing such gross ignorance of the actual vergence state, whilst maintaining an acute awareness in changes in vergence, seems hard to sustain.

Fifth, the argument that the visual system can extract absolute distance (theta) from changes in vergence (delta theta) has been repeatedly put to us as a way of preserving vergence as an important absolute distance cue in spite of our experimental results. But the suggestion that vergence is blind to gradual changes may be just as damaging. If we maintain that vergence is blind to gradual changes, but it later turns out (as one would expect) that subjects are no less accurate in judging near distances in full-cue conditions when the distance of the target is gradually manipulated (e.g. by pointing to objects in full-cue conditions), then advocates of this account would have to concede that there is no benefit from having vergence as an absolute distance cue when reaching for objects; the very scenario where vergence is thought to be at its most important (Loftus, Servos, Goodale, Mendarozqueta, & Mon-Williams, 2004; Melmoth & Grant, 2006; Melmoth, Storoni, Todd, Finlay, & Grant, 2007). So this account risks replacing the ineffectiveness of vergence under my account with the redundancy of vergence under their account.

Conclusion

Vergence is considered to be one of our most important absolute distance cues. But vergence has never been tested as an absolute distance cue divorced from obvious confounding cues such as binocular disparity. In this article we control for these confounding cues for the first time by gradually manipulating vergence, and find that observers fail to accurately judge distance from vergence. We consider a number of different interpretations of these results. Whilst ad-hoc reinterpretations of vergence as blind to gradual changes, or reliant on delta theta rather than theta, cannot be definitively ruled out, we argue that the most principled response to these results is to question the effectiveness of vergence as an absolute distance cue. Given other absolute distance cues (such as motion parallax and vertical disparities) are limited in application, this poses a real challenge to our contemporary understanding of visual scale.

6. Vergence as a Motion in Depth Cue

The conclusion that vergence is an ineffective visual depth cue would explain the absence of motion-in-depth in Erkelens & Collewijn (1985a; 1985b), and would help us to unify the distance and motion-in-depth literature. However, the finding in Erkelens & Collewijn (1985a; 1985b) has been heavily revised over the last 30 years, and there is now significant evidence that vergence contributes to motion-in-depth. Can my thesis explain these observations?

I would argue that it can. The raft of papers that emerged in 2006-2010 only sought to prove that vergence made *a* contribution (as opposed to *no* contribution) to our motion-in-depth estimates. In the words of Welchman et al. (2009), it was "surprising that the view that vergence velocity signals do not support perceptual estimation has remained largely unchallenged for over twenty years." By contrast, I am happy to admit that vergence makes a modest contribution to our motion-in-depth estimates. But my question is whether this contribution is *visual* (vergence affects our visual experience), or merely *kinesthetic* (an independent physical sensation)?

To give an illustrative example, consider how *visual* (optic flow) and *kinesthetic* (vestibular) cues interact in the experience of illusory self-motion. Vestibular cues are not necessary for illusory self-motion (Fischer & Kornmüller, 1930), e.g. when we see a neighbouring train pull away in a railway station whilst we are stationary. Nonetheless, the addition of vestibular acceleration can contribute to the illusion of self-motion (Ash et al., 2011). But, and this is the crucial point, there is no suggestion that this contribution has to be *visual*: It is not as if the optic flow has to be seen as flowing faster in order to incorporate this vestibular cue into our determination of self-motion. Instead, the visual (optic flow) and the non-visual (vestibular) cues merely feed into a common post-perceptual determination. Indeed, as Ash, Palmisano, & Kim (2011) observe, there is no 'mandatory fusion' between the *visual* (optic flow) and *kinesthetic* (vestibular) cues, which is often taken to be the litmus test of a truly perceptual effect: see Hillis, Ernst, Banks, & Landy (2002).

But if this holds true for self-motion, why can't it hold true for object motion? After all, the two scenarios are symmetrical: the same *visual* cue (optic flow, which is identical whether it is caused by self-motion or object motion) is confirmed by a *kinesthetic* cue (*vestibular* in the case of self-motion, *vergence* in the case of object motion) by feeding into a common post-perceptual determination. Let us consider how this hypothesis fits with the empirical literature:

1. The first argument against vergence as a purely *kinesthetic* cue is that the same change in angular rotation would produce the same apparent motion, even though the motion-in-depth for 0° to 5° (from the horizon to 73cm) is much greater than the motion-in-depth for 5° to 10° (from the 73cm to 37cm). But Lugtigheid, Brenner, & Welchman (2011) find that motion-in-depth from vergence isn't scaled for absolute distance, and does in fact simply rely upon the change in the angular rotation of the eyes.

2. The second argument against a purely *kinesthetic* account of vergence is that such a cue would still be present in those instances where we don't see motion-in-depth (for instance a 30° x 30° stereogram oscillating in depth: Erkelens & Collewijn, 1985a; 1985b), so why do we fail to detect motion in this context? The problem is starkly illustrated by Welchman, Harris, & Brenner (2009), who confirm that subjects can use vergence to judge the direction of motion of a 0.1° x 0.1° square, but fail for a 22° x 17° field of triangles: "This confirms previous findings that extra-retinal signals alone do not inform perceptual estimates of motion direction for large-field stimuli that do not change in retinal size (Erkelens & Collewijn, 1985b; Regan et al., 1986)." As Welchman et al. explain, they used a field of triangles to "make sure that motion-in-depth of the large background was imperceptible." Both Welchman et al. and I agree that vergence is not functioning as a *visual* cue in this context.

But unlike Welchman et al., I argue that vergence must be acting as a non-visual *kinesthetic* cue in this context, since (contrary to the above quotation) subjects performed quite well when detecting the motion-in-depth of the 22° x 17° field of triangles. When we look at Welchman et al.'s data (their fig.3a, 0ms), we find thresholds for detecting the direction of motion for a 22° x

17° field of triangles are around 0.7°/s (12cm/s), which is (a) only double the threshold for a 0.1° x 0.1° square, and (b) an order of magnitude lower than the 6°/s to 13.5°/s tested by Erkelens & Collewijn (1985b) for which subjects are still *visually blind* to motion. So subjects are unable to see the motion, but they are nonetheless able to reliably judge its direction. This is only a paradox if we are committed to thinking that all motion-in-depth cues must operate *visually*. They clearly don't in the context of illusory self-motion (with vestibular cues), so why think they must in the context of object motion?

But why has this apparent paradox been overlooked? Because Welchman et al. (2009)'s apparatus had a theoretical maximum of 2.2°/s (50cm/s), thresholds had to be less than 0.4°/s (7.5cm/s) to reach statistical significance. Since the threshold for the 22° x 17° field of triangles was ≈ 0.7°/s (12cm/s) it didn't reach statistical significance, but the threshold for the 0.1° x 0.1° square (after 300ms) was ≈ 0.3°/s (6cm/s) so it did (cf. the performance after 100ms and 200ms). In the context of the 22° x 17° field of triangles Welchman et al. interpret the absence of statistical significance as confirmation of the null hypothesis. But this gives a misleading impression of the modest differences between their conditions (see their fig.3a). Specifically, it eradicates the performance in the 22° x 17° field of triangles condition, rather than explaining it. The only explanation for this performance, if we are committed to the idea that subjects are *blind* to the motion-in-depth of large-field stimuli, is that subjects were aware of their eye movements by means of a non-visual *kinesthetic* cue: even if subjects couldn't see the motion, they could feel the rotation of their eyes.

3. The third argument against a purely *kinesthetic* account of vergence is that although *visual* motion-in-depth was absent in Erkelens & Collewijn (1985a; 1985b), it was later found to be present in Regan et al. (1986) when subjects viewed a single dot moving in depth.

But I would argue that this is actually the strongest argument in favour of vergence as a purely *kinesthetic* cue: If vergence were an effective *visual* cue to motion-in-depth, then we would expect a vivid impression of motion as the eyes tracked the single dot. And yet this isn't the case:

Regan et al. (1986) describe the motion-in-depth as 'weak', Harris (2006) "confirms that Z-direction motion is indeed difficult to detect", and the subjects in Howard (2008) report only 3.7cm gain for a 37.5cm change in vergence. Indeed, the motion-in-depth was so poor that Harris (2006) concluded that "the visual system appears not to use vergence eye movement signals in the perception of the Z-component of 3-D motion." This is despite the fact that we know from Brenner & van Damme (1998) that vergence contributes to effective *judgements* of relative depth, so whatever contribution vergence is making, it doesn't appear to be *visual*.

Furthermore, the modest motion-in-depth that is reported can be explained by retinal slip. Welchman et al. (2009) exclude the possibility that retinal slip is *solely* responsible for our motion-in-depth estimates. They move a stimulus (a dot with 28° background) at +/- 0.57° over 1.2 seconds (equivalent to 7cm towards or 9cm away from the observer at a 50cm viewing distance), and then the background disappears and the dot either changes speed / direction (inducing retinal slip) or maintains a constant speed, and participants have to judge the direction of the dot. Welchman et al. (2009) work on the assumption that participants are blind to the motion of the dot + background (which I've criticised above), and then compare whether the dot was judged to be approaching or receding with the participants' eye tracking data to determine whether vergence had any effect over and above retinal slip.

[Removed due to copyright]

Figure 36. Results for Experiment 2 in Welchman et al. (2009).

To understand their results in Fig.36, we have to comparing the four quadrants. Compare the top-right (vergence is approaching and retinal slip suggests approaching – i.e. vergence is struggling to catch up with an approaching target) with the bottom left (vergence is receding and retinal slip suggests receding – i.e. vergence is struggling to catch up with a receding target). There's clearly more red points (judgements the stimulus is approaching) in the top right quadrant than in the bottom left quadrant, and that is what we'd expect. Now, in order to work out how much those two signals (vergence and retinal slip) are contributing to this difference between the top right and bottom left quadrants, we compare the following quadrants. For retinal slip, we compare the left quandrants with the right quadrants, and see a clear shift in the number of red vs blue points. But the key point that Welchman et al. (2009) are making is that if we compare the top vs bottom quadrants we also see a change in the distribution of red vs blue points, and this corresponds to the change in vergence direction influencing the judgement of target direction.

Putting to one side concerns about the accuracy of eye tracking (which would add noise to the signal), another concern is that the vergence motion was present for 1.2 seconds (and in the judged direction for 0.45s), whilst the retinal slip was present for just 0.15s. But taking these results at face value, what they show is that vergence makes a marginal contribution to direction judgements, which is entirely consistent with the kinesthetic account. However, these results demonstrate that subjects perform substantially worse in detecting the motion of a dot when they track its motion with vergence rather than simply keeping their eyes fixed. This suggests that whatever contribution vergence does make to the discrimination of motion-in-depth, it is substantially less than the retinal slip it replaces. In this regard, vergence also appears to supress motion-in-depth rather than being an effective cue to it.

4. The fourth argument in favour of vergence as a *visual* cue is that when motion-in-depth is produced by 'looming' (a sudden increase in angular size), the addition of a consistent vergence signal appears to accentuate the motion (Heuer, 1987; Brenner, Van Den Berg, & Van Damme, 1996; 'texture alone' in Howard, Fujii, & Allison, 2014, fig.4; cf. 'texture alone' in Howard et al., 2014, fig.5). However:

First, the size of this effect is relatively modest, if it is present at all: Brenner, Van Den Berg, & Van Damme (1996) find only "a slightly lower velocity" from the absence of vergence, and in the one instance Howard, Fujii, & Allison (2014) do find an improvement, the gain is only about 10% (from 20cm of motion-in-depth to 22cm). There is no reason why an effect of this magnitude couldn't be accounted for by a purely *kinesthetic* cue.

Second, there is virtually no evidence that vergence in the wrong direction has the opposite effect (Heuer, 1987; 'texture alone' in Howard et al., 2014, fig.5; cf. Swanston & Gogel, 1986), even though, as we have already noted, such reduced performance ('mandatory fusion') is the litmus-test for a truly perceptual effect: see Hillis et al. (2002). Indeed, Heuer (1987) suggests that subjects can consciously choose which of the two contradictory signals – vergence or looming – to attend to. So just as with illusory self-motion (Ash, Palmisano, & Kim, 2011), we find no 'mandatory fusion' between the *visual* and *kinesthetic* cues. Indeed, this appears to be a trend when *visual* cues are integrated with *physical sensations*, be they *kinesthetic* or (in the case of Hillis et al., 2002) *haptic*.

5. Fifth, Lugtigheid et al. (2011) find vergence can modulate motion-in-depth from relative disparity by up to 60%. But this impressive figure needs to be put into context: (1) The relative disparity is moving towards us at roughly 3.7cm/s. (2) Adding a convergent eye movement of 13.6cm/s adds an impression of forward motion of about 1cm/s, so the relative disparity in the stimulus only has to travel at 2.7cm/s to achieve the same motion. (3) Adding a divergent eye movement of 13.6cm/s seems to wipe off forward motion by about 1.3cm/s, so the relative disparity has to travel at 5cm/s to achieve the same forward motion. So, as with vergence's contribution to looming, the gain is an order of magnitude less than the vergence signal itself. And again, there is no reason why this effect could not be accounted for by a *kinesthetic* cue.

6. Sixth, Nefs & Harris (2008) find vergence is important for the induced motion illusion, where a fixed target appears to move in the opposite direction from an actually moving target. But an alternative explanation for this finding is that induced motion in depth only occurs when (a) you are not fixated on the static target, and (b) there is relative disparity between the static and moving target. The only experimental condition in Nefs & Harris (2008) that satisfied those two conditions necessarily involves changing vergence.

In conclusion, over the last decade the motion-in-depth literature has been driven by the assumption that "the distance of a stationary object can be judged on the basis of vergence alone. So why was motion-in-depth not produced by changing vergence?" (Howard, 2012). Once we realise that stationary vergence isn't an effective distance cue, the impetus for this literature is removed. The last decade of effort has seen the accumulation of studies with very modest gains in motion-in-depth from vergence. This is consistent with (and indeed, would appear to confirm) the notion that changing vergence merely functions as a *kinesthetic* cue.

7. Vergence Scaling

Vergence is supposed to be fundamental to two scaling mechanisms: (1) size constancy, and (2) depth (or 3D shape) constancy. These two scaling mechanisms have to be reassessed in light of vergence's ineffectiveness as a distance cue:

1. Size Constancy: When we scale a visual scene, we are trying to differentiate a small object up close from a large object far away. In this sense, size and distance are inextricably linked. We've known since Ptolemy (c.160 AD) that the visual angle of the stimulus does not determine this question, and that the visual system has to rely on distance information to specify the size as well and distance of the stimulus. Some of the earliest studies of vergence as a distance cue (Meyer, 1842; Wheatstone, 1852) suggested that vergence scales the size of the stimulus, because the stimulus appears to shrink as vergence is increased. But if vergence is an ineffective distance cue, then how are we to explain this 'vergence micropsia'?

One suggestion is that vergence is a distance cue for size and disparity, but not for distance itself. Ono & Comerford (1977) consider the possibility that "the question whether convergence serves as a cue to distance can be divided into two parts: (a) does oculomotor adjustment provide distance information for the visual system and (b) is the distance information provided by oculomotor adjustment, if any, used to make distance judgments?" Bishop (1989) makes a similar point. But this would be a surprising conclusion to have to come to. First, it seems paradoxical that distance should be a cue for size and disparity, but not distance itself. Second, as we have just observed, size and distance are inextricably linked so far as 'scale' is concerned: we want to differentiate a small object up close from a large object far away, yet what we'd have is one parameter (size) changing without the other (distance).

The alternative is to explore whether vergence micropsia is an artefact of stimulus presentation. We can begin to see why it might be if we cross-fuse the two coins in Fig.28:

Figure 37. Two coins to cross-fuse as a demonstration of vergence micropsia.

What we notice are two idiosyncrasies of vergence micropsia: First, vergence micropsia shrinks the central fused coin, but appears to leave the two monocular flankers unaffected (this works in the opposite direction when the coins are parallel-fused). Second, this asymmetry in size between the central fused coin and the monocular flankers is only accentuated as we reduce our distance to the screen.

What might explain this effect? Consider how the two coins are viewed when your eyes are not crossed: The distance of the left coin from the right eye is slightly greater than the distance of the right coin from the right eye. Each coin is made up of a larger retinal image (right eye in the case of the right coin, left eye in the case of the left coin) and a smaller retinal image (left eye in the case of the right coin, right eye in the case of the left coin). But when we cross-fuse, what happens? The central coin is made up to the two smaller retinal images, whilst the two monocular flankers are made up of the remaining larger retinal images, leading to the asymmetry that we see between the central fused coin and its monocular flankers.

The same concern applies when vergence is varied in a stereoscope by increasing the separation between the stimuli for the left and right eye: First, as the stimulus for each eye travels increasingly in the opposite direction, the distance of the stimulus from the eye increases, introducing small reductions in x-axis and y-axis size. Second, as the stimuli are presented on a fronto-parallel display, when the distance from the eye increases the stimuli are also rotated relative to the eye, introducing another small reduction in x-axis size. Third, this rotation doesn't just reduce the x-axis size, it also distorts the shape of the object. We simply don't know the x- and y-axis size of the eventual fused percept is affected when the visual system has to reconcile two inconsistently distorted shapes.

One option is to correct for these distortions and see if vergence micropsia nonetheless persists. This is the focus on Experiment 3.

The alternative is to fix the retinal image with after-images viewed in darkness. There are reports of vergence micropsia when subjects view an after-image of their hand and then move their real hand backwards and forwards in space (or close variations of this paradigm; see: Taylor, 1941; Gregory, Wallace, & Campbell, 1959; Morrison & Whiteside, 1984; Suzuki, 1986; Carey & Allan, 1996; Mon-Williams et al., 1997; Bross, 2000; Ramsay, Carey, & Jackson, 2007; Sperandio, Kaderali, Chouinard, Frey, & Goodale, 2013; Zenkin & Petrov, 2015).

2. 3D Shape Constancy: Mis-scaled binocular disparities are liable to distort both the 3D shape of objects and the 3D shape of the visual field itself.

a. Shape of Objects: Whilst angular size reduces in proportion to the distance of the object, binocular disparity reduces in proportion to distance of the object squared. What this means is that the z-axis depth of an object drops off more rapidly with distance than its x-axis and y-axis size. So without the absolute distance being used to rescale the fall-off in binocular disparity to the fall-off in angular size, binocular disparity will not preserve the 3D shape of objects with distance.

Whilst this coheres with our impression of real-world scenes extending off into the far distance (the z-axis depth of distant buildings appears to fall off more quickly than their size, leaving them, in Vishwanath, 2010's terms, 'almost *pictorial*'), for a couple of decades it was believed that within vergence's effective range (up to 2m) the visual system not only compensated for this drop-off in z-axis depth relative to x-axis and y-axis size (i.e. preserved the object's 3D shape), but actually recovered the object's absolute metric depth: see Wallach & Zuckerman (1963); Fried (1973); Ono & Comerford (1977); Wallach, Gillam, & Cardillo (1979).

This consensus was eroded by Foley (1980) and Johnston (1991), the latter finding that the 3D shape of a cylinder defined by disparity was distorted with distance (at 53.5cm it was elongated towards the viewer, at 107cm it was veridical, and at 214cm it was compressed). This distortion of 3D shape with viewing distance has been confirmed by Glennerster, Rogers, & Bradshaw (1996); Bradshaw, Parton, & Glennerster (2000); and Scarfe & Hibbard (2013).

There have been three responses to this shape inconstancy: (1) Quarantine it to specific tasks / contexts (Glennerster et al., 1996; Bradshaw et al., 2000). (2) Eradicate it by introducing other depth cues (Scarfe & Hibbard, 2013; Guan & Banks, 2016). (3) Simply admit that the 3D shape of objects can be distorted with viewing distance (Morgan, 1989; Johnston, 1991; Campagnoli et al., 2017).

But the admission that 3D shape can be distorted with viewing distance only addresses the fact that shape constancy is not veridical. I still need to explain why vergence appears to make some contribution to shape constancy. The typical suggestion is that vergence is scaling disparity, but using a non-veridical distance estimate (Foley, 1980; Johnston, 1991; Campagnoli et al., 2017). But there is another phenomenon that needs explaining: When we increase our distance from a stereo-image, by walking away or moving our head back, the scene appears to expand in z-axis depth relative to the x-axis and y-axis, and compress when we approach it. First, the geometry of the scene is vividly distorted as we move our head back and forth, suggesting (vs. Scarfe & Hibbard, 2013) that the addition of other depth cues does not eradicate such distortions. Second, the scaling of the depth in the scene is the opposite of the relationship we would expect: as vergence reduces, the depth in the scene increases. This is in conflict with the literature from Helmholtz (1866) on telestereoscopic viewing, through to Regan et al. (1986), Cumming, Johnston, & Parker (1991), and Johnston (1991), who all suggest that increasing vergence increases the perceived depth of the stimulus. As we discussed above, Johnston (1991) found that the depth of a RDS of a cylinder was found to be accentuated at near distances, and reduced at far distances. Our own hypothesis is that Johnston (1991) failed to appropriately control for angular size reductions (see our explanation of z-scaling from x and y-axis scaling in the next paragraph), but there's no question that this is the orthodox account in the literature.

Wallach et al. (1979) suggest that the failure of 3D shape constancy (or 'depth constancy') when moving your head away from a stereogram is caused by the fact that for a fixed-disparity image (such as a stereo-image) disparity falls off linearly with distance, so depth scaling (which responds to a fall-off of disparity with distance squared) necessarily overcompensates. But this doesn't address the inconsistency with the classic claim of the literature just discussed that when disparity is fixed, and vergence is increased, this is supposed to increase, not decrease, the depth from disparity. We propose a way of reconciling these otherwise inconsistent observations. What is common to both scenarios is a reduction in x-axis and y-axis size, due to an increase in physical distance (in Wallach et al., 1979) or vergence micropsia (in Cumming et al., 1991). Perhaps the increased impression of depth in both of these contexts is simply a function of constant retinal disparity seen in relation to a reduction in x-axis and y-axis size?

b. Shape of the Visual Field: The binocular disparity of a fronto-parallel surface varies with eye rotation, and therefore vergence distance. But the shape of fronto-parallels do not appear to change with distance, which suggests a scaling mechanism. Fronto-parallel constancy was originally seen as less effective than shape constancy (see Ono & Comerford, 1977's discussion of Luneburg, 1947). Then they were regarded as equally non-veridical (Foley, 1980). Now this position has reversed, and it is fronto-parallel constancy which is regarded as close to veridical, and shape constancy which is regarded as comparatively ineffective (Rogers, 2006).

8. Experiment 3 (Vergence Size Constancy)

Distance estimates derived from vergence are thought to be a key contributor to 'size constancy', the process of scaling the retinal image to maintain a constant impression of physical size despite changes in the retinal image size caused by changes in distance. Emmert's Law (Emmert, 1881; Darwin, 1786) states that the apparent size of a retinal after-image is perfectly corelated with the distance of fixation. Size constancy from pictorial cues alone is merely partial (10%-45%: Murray et al., 2006; Leibowitz et al., 1969), leading to the suggestion that size constancy is near perfect in real-world viewing (Emmert's Law) "because the available monocular cues are completely congruent with more reliable sources of distance information such as vergence, accommodation, stereopsis, and motion parallax" (Sperandio et al., 2012). Combe & Wexler, (2010) refer to "the common notion that size constancy emerges as a result of retinal and vergence processing alone" (although suggesting a role for motion parallax), whilst Sperandio & Chouinard (2015) suggest that vergence (eye rotation) and accommodation (focus state of the eye) "are not only powerful but they are also essential when viewing conditions are reduced or restricted."

Evidence for this conclusion comes from 'vergence micropsia', the finding (in four specific contexts) that changing the angular rotation of the eyes affects the perceived size of objects:

1. Wallpaper Illusion: If you cross your eyes whilst looking at a recurring wallpaper pattern, the pattern appears smaller and closer. This observation provided the earliest evidence supporting vergence micropsia (Smith, 1738; Priestley, 1772; Goethe, 1810; Meyer, 1842, 1852; Brewster, 1844; Locke, 1849; Lie, 1965; Ono et al., 1971; Kohly & Ono, 2002; see Howard, 2012).

2. Stereoscopic Viewing: Haploscopes present separate images to the two eyes, and therefore enable the vergence angle to be changed whilst maintaining the same image to each eye. If vergence is increased, the perceived image appears to shrink, even though its physical size remains fixed (Wheatstone, 1852; Helmholtz, 1866, p.313; Judd, 1897; Frank, 1930; Hermans, 1937, 1954; Locke, 1938; Adams, 1955; Von Holst, 1955a, 1955b, 1957; Heinemann et al., 1959;

Gogel, 1962; Biersdorf et al., 1963; Wallach & Zuckerman, 1963; McCready, 1965; Leibowitz & Moore, 1966; Leibowitz et al., 1972; Komoda & Ono, 1974; Regan et al., 1986; Enright, 1989).

3. Telestereoscopic Viewing: If we use mirrors to artificially increase the distance between the viewpoints of the two eyes, the world appears miniaturised: "it will seem as if the observer were not looking at the natural landscape itself, but a very exquisite and exact model of it, reduced in scale" (Helmholtz, 1866, p.312). This effect has been attributed to vergence by Helmholtz (1857; 1858; 1866, p.310) and Rogers (2009; 2011), and has been well studied in military research where helicopter pilots view the world through cameras with increased separation (Newman & Ostler, 2009; Stuart et al., 2009; and Priot et al., 2010; 2011; 2012; 2018; cf. Linton, 2018).

4. Taylor Illusion: If you flash a bright enough light in a dark room, you leave a temporary after-image on the retina of whatever object was illuminated. If you make an after-image of your hand, and then in complete darkness move your physical hand closer, the after-image of your hand appears to shrink even though it doesn't change size on the retina (the 'Taylor illusion'; Taylor, 1941). The best current explanation for the Taylor illusion is that it is due (Taylor, 1941; Morrison & Whiteside, 1984; Mon-Williams et al., 1997) or almost entirely due (Sperandio et al., 2013) to the increase in vergence as the eyes track the invisible physical hand coming closer (for further discussions of the Taylor illusion see Gregory et al., 1959; Carey & Allan, 1996; Bross, 2000; Ramsay et al., 2007; Faivre et al., 2017a; and for vergence scaling of after-images see Urist, 1959; Suzuki, 1986; Lou, 2007; Zenkin & Petrov, 2015; and Millard et al., 2020 on the vergence scaling of after-images). The reason why, is that when Sperandio et al. (2013) asked subjects to track an LED moving in depth in the opposite direction from the hand, they found the after-image changed size according to vergence, not the hand movement, and the size of the size change was almost as large as when both the hand and vergence were moving in the same direction.

Surveying the literature, two things are striking: First, to our knowledge, there has never been a report of a failure of vergence micropsia at near distances (25-50cm). For instance, when Regan et al. (1986) found that vergence as a motion in depth cue could be vetoed by using a full-field stimulus, they nonetheless reported that even then "apparent size changes as about threefold when convergence changed from about 0 deg to 25 deg.", commenting that "Changes in size and depth produced by ocular vergence changes are well known". Second, the after-image literature appears to suggest something close to perfect size constancy for vergence in reaching space (25-50cm). First, because the near-perfect size constancy of the Taylor illusion (Bross, 2000; Ramsay et al., 2007; Sperandio et al., 2013) can be almost entirely attributed to vergence (Sperandio et al., 2013). Second, because size constancy for after-images also appears close to perfect for 25-50cm when vergence is the only distance cue. Apparent size doubled for the representative subject in Sperandio et al., (2013) (incongruent condition) from 3.3cm at 25cm (suggested by the $y = -0.61x + 3.3$ line of best fit) to 6.3cm at 50cm (average of size estimates after a >3° vergence eye movement) (my analysis of data from their Fig.5 using WebPlotDigitizer 4.2; Marin et al., 2017).

Vergence size constancy is therefore regarded as a fundamental aspect of visual scale processing. However, we believe vergence size constancy should be re-evaluated for two reasons:

First, our recent work suggests that vergence is an ineffective absolute distance cue once confounding cues have been controlled for. Participants are unable to use vergence to judge absolute distance (Linton, 2018; in press), and we are reluctant to embrace the suggestion (discussed by Ono & Comerford, 1977 and Bishop, 1989) that vergence may be effective distance cue for size constancy, but not for conscious absolute distance judgements.

Second, to our knowledge, all previous tests of vergence size constancy have introduced confounding cues, and provided participants with additional distance information. This is because vergence can only be driven experimentally in one of two ways. Either participants track a visual object moving in depth (such as an LED: Mon-Williams et al., 1997; Sperandio et al., 2013), in which case they are informed about the change in distance by the visible motion of the object moving in depth. Or subjects track their own hand (as in the Taylor illusion), but again this gives them information about the distance change from proprioception. So the purpose of our experiment was to test vergence size constancy, in a context where it is supposed to be effective (vergence size change over 5 seconds from 25cm to 50cm; Sperandio et al., 2013), but in a way that doesn't introduce subjective knowledge about the distance change to the participants.

Methods

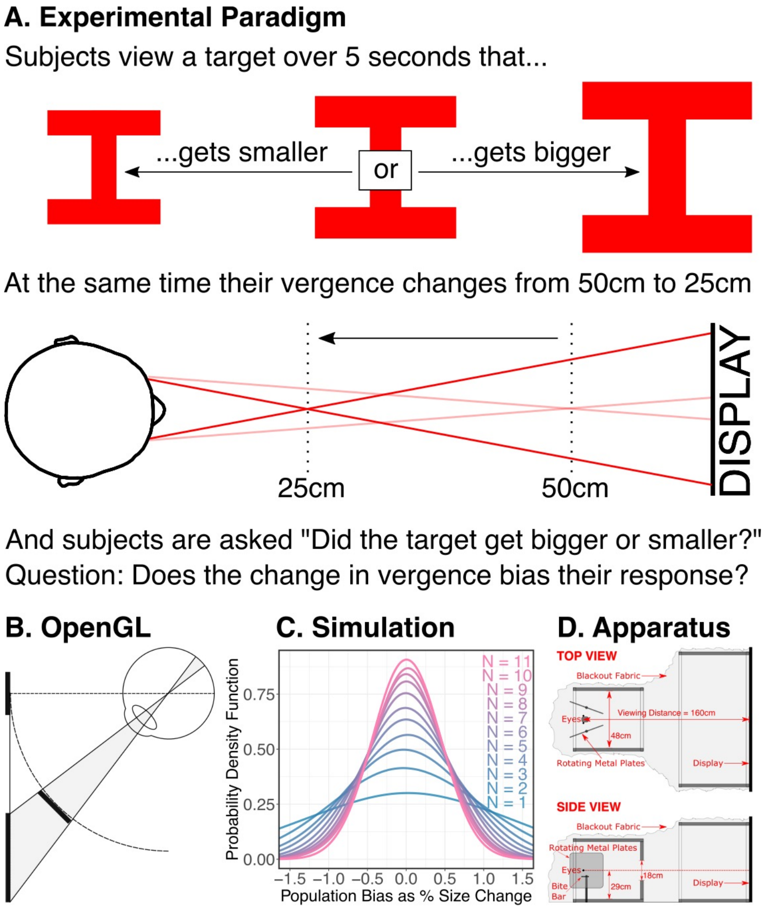

Figure 38. Experimental Methods for Experiment 3. A. Outline of the experimental paradigm. B. To maintain the correct retinal image with eye rotations, we have to simulate a virtual target moving in an arc with a constant distance and orientation to the eye, and project this image onto the fronto-parallel display. C. We simulated the experiment 10,000 times in Quest+ (bias = 0, detection threshold = 5%, lapse rate = 2%), and modelled how increasing the number of participants would improve the accuracy of our hierarchical Bayesian estimate of the bias (true bias = 0), in order to determine we need N = 5+ to rule out vergence size constancy > 1.5%. D. Diagram of apparatus.

This posed a complex technical challenge which we solved in five ways:

1. First, in order to drive vergence without providing subjective distance information, we used a visual stimulus that (unlike an LED) provided 'sub-threshold' binocular disparities: binocular disparities that are visible to the participant's visual system (in order to drive vergence), but subjectively invisible to the participant themselves. This we achieved with a 3° target moving in depth from 50cm to 25cm over 5 seconds. In reality, the target consisted of two targets on a display: a left-hand target that only the right eye could see, and a right-hand target than only the left eye could see. And by increasing the separation between the targets, we could increase the participant's vergence angle (Fig.38A). The rotating metal plates in Fig.38D ensured that each eye only saw the appropriate target, and participants were asked to confirm that they saw a single fused target when they opened both eyes, and to report if the target ever went double.

2. Second, although we manipulate vergence by increasing the lateral separation between the targets on the display in Fig.38A, as Fig.38B demonstrates, if we just want to isolate vergence as a size constancy cue, we need to present a constant retinal image to the eye, and this won't be achieved by simply moving the target laterally on the display. Instead, what we need to present is a target that maintains a constant radius and orientation to the eye as vergence is increased. This is the ideal, and we can simulate it on the physical display by using OpenGL to 'project' this simulated target onto the physical display behind it to be viewed by the observer (Figure 39) (technically speaking, we set the camera frustum in OpenGL to be the nodal point of the eye, and use an asymmetric frustum so that the far clipping plane matches the distance and dimensions of the display). A bite bar was used to ensure that the nodal point of the eye remained fixed during the experiment (Figure 39), and the difference between the nodal point and the center of rotation of the eye was intentionally ignored (cf. Linton, 2019; Konrad et al., 2019).

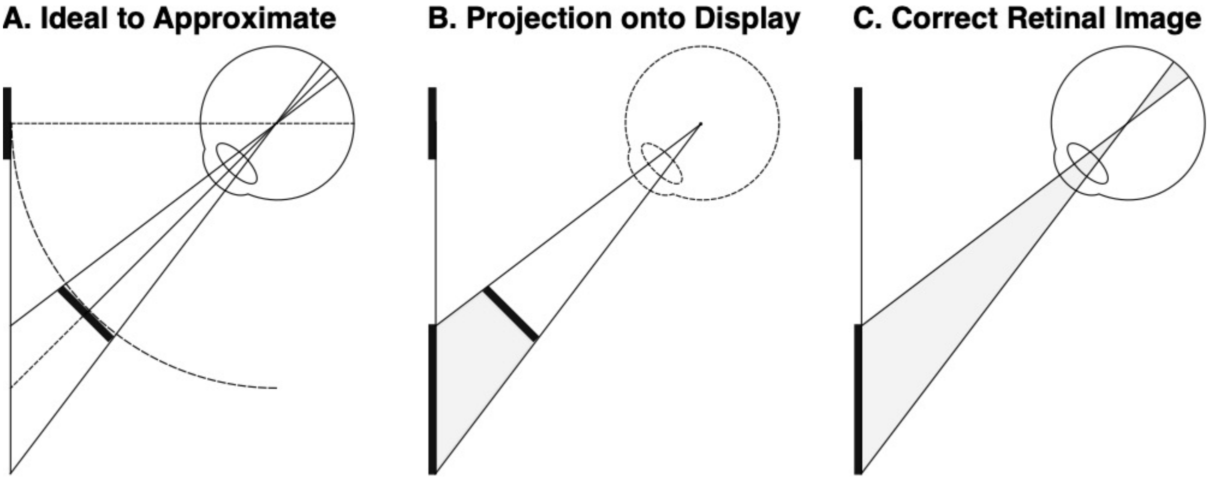

Figure 39. Illustration of the principle behind OpenGL projection in Experiment 3.

3. Third, another challenge of this display is that it requires the eyes to focus (or 'accommodate') at the distance of the display (160cm), whilst vergence (the angular rotation of the eyes) is at 25-50cm. This doesn't happen in normal viewing conditions, and too much vergence-accommodation conflict can lead to the target going blurry or double. To solve this problem we had an optometrist fit each participant with contact lenses (based on the participant's valid UK prescription) so that the optical distance of the display was 33cm even though its physical distance was 160cm. This ensured a maximum of +/– 1 dioptres worth of vergence-accommodation conflict, well within the zone of 'clear single binocular vision' (Hoffman et al., 2008). Some of the most dramatic reports of vergence micropsia have been in the presence of large vergence-accommodation conflicts (e.g. 6.5 dioptres in Regan et al., 1986), so the presence of +/– 1 dioptre should not be objectionable. Contact lenses were used in preference to trial lenses because trial lenses risked introducing off-axis distortions and optical micropsia.

4. Fourth, we wanted the target to be presented in perfect darkness to exclude any remaining visual cues. An interesting finding from piloting was that the usual technique (having participants view a CRTs through neutral density filters) wasn't effective in eradicating residual luminance from the display (it degraded the target before residual luminance was completely eradicated). Instead, we achieved this ideal in four ways: First, we used an OLED display (LG

OLED55C7V) which unlike normal displays does not produce residual luminance for black pixels. Second, subjects wore a mask to block out any residual light, which had red eye filters through which the red stimuli were viewed (blocking out 100% green and ~90% blue light). Third, subjects viewed the stimuli through a narrow (17°) viewing window of 48cm x 18cm at a distance of 60cm. Fourth, the whole apparatus was covered by blackout fabric, and before the experiment began subjects pulled a hood of blackout fabric over their heads and the external lights were turned off.

5. Fifth, rather than ask participants to match their visual experience to a visible chart (Bross, 2000; Lou, 2007; Sperandio et al., 2013) or to a memorised chart (Ramsay et al., 2007), or ask for conscious judgements of the physical size of the after-image (Mon-Williams et al., 1997), or its % size change (Carey & Allan, 1996), we built size change estimation into the stimulus itself (something that cannot be done with an after-image). We increase or decrease the physical size of the target on each trial by between –20% and +20%, and ask participants to make a forced choice ("did the target get bigger or smaller?"). The question is whether the vergence change biases the response? If there is no vergence size constancy, then we would expect the answer to be 'no'. Participants will simply be at chance in determining whether there is a size change when there is no physical size change. By contrast, if participants experience vergence micropsia, then the degree of the bias (i.e. the degree of physical size increase we have to introduce before participants are at chance) will indicate just how large the vergence micropsia effect is.

We used a four-parameter maximum likelihood model (Quest+: Watson, 2017; Brainard, 2017) to estimate when participants were at chance. Participants completed 200 trials (10 sets of 20 trials), and on each trial Quest+ tested the size increase or decrease that would maximally improve our estimate of participant bias. In piloting, we found that the author could not detect size changes smaller than 1.5%, so if vergence size constancy changes perceived size by less than 1.5% it can be dismissed as unimportant. We simulated the experiment 10,000 times in Quest+ (bias = 0, detection threshold = 5%, lapse rate = 2%), and fit a hierarchical Bayesian model (discussed below) to model how increasing the number of participants would improve the accuracy of our estimate of the bias (true bias = 0) (Fig.1C). We found that with 5 or more observers we could rule out any vergence size constancy greater than 1.5%. The visual stimuli were presented in MATLAB 2019a (MathWorks) using OpenGL in PsychToolBox 3.0.15 (Kleiner et al., 2007) and the code for running the experiment is openly available: https://osf.io/5nwaz/

11 observers (8 female, 3 male; age ranges 20-34, average age 24.5) participated in the study: the author and 10 participants recruited using an online advertisement (13 were originally recruited, but 1 was excluded because they could not fuse the target, and 2 were excluded because they could not get clear vision with the contact lenses). All participants were screened to ensure accommodation was within normal bounds for age (tested with a RAF near-point rule), vergence within normal bounds (18D or above on a Clement Clarke prism bar), and stereoacuity within normal bounds (60 arc secs or less on a TNO stereo test). The author's participation was required to (a) confirm Quest+ mirrored the pilot data, and (b) provide a criterion for the minimum effect size. All other subjects were naïve as to the purpose of the experiment, and paid £15/hr for 3 hours. The study was approved by the School of Health Sciences Research Ethics Committee at City, University of London in accordance with the Declaration of Helsinki.

Results

Fig.40A illustrates a summary of the results, plotted against what we would expect according to a number of hypotheses. As we have already discussed, if vergence has no effect on perceived size, then participants should be at chance when we don't introduce any physical size change, which is exactly what we found. By contrast, had participants experienced full size constancy (equivalent to the representative participant in Sperandio et al., 2013), then we would have had to increase the size of the target by 100% to counter the reduction in distance by 50% (because in normal viewing conditions, halving the distance leads to a doubling of the retinal image).

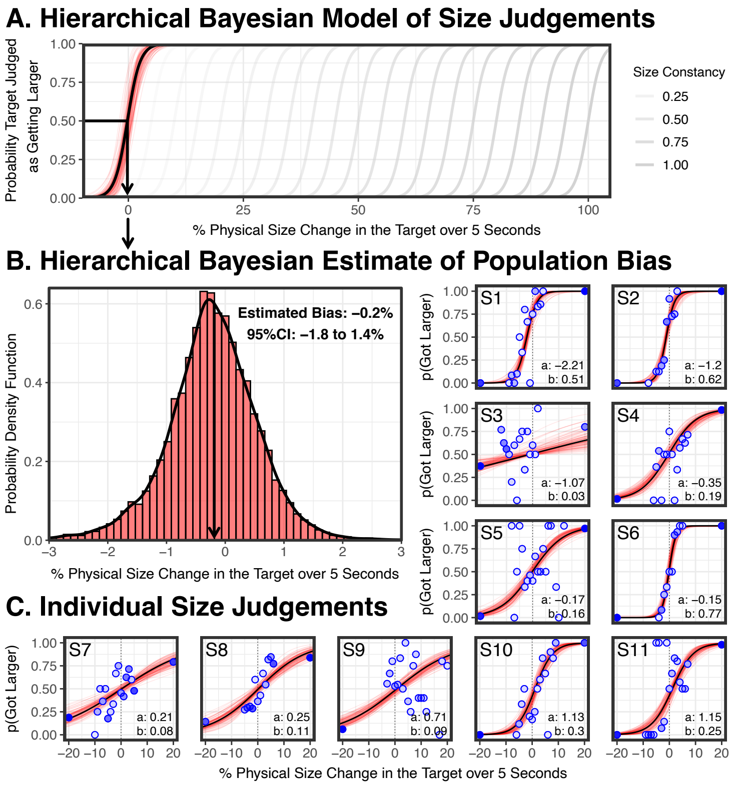

Figure 40. Results from Experiment 3. A. Hierarchical Bayesian model of the population psychometric function in black (based on 15,000 posterior estimates, 100 representative posterior estimates in red), plotted against predictions for various degrees of vergence size constancy effect sizes (in grey). B. Probability density function of 15,000 posterior estimates of the population bias, with a non-significant bias of –0.2%. Fig.2C. Bayesian psychometric functions in black fitted to individual subject results (based on 15,000 posterior estimates, 100 representative posterior estimates in red). Blue dots indicating the size changes tested by Quest+ (with darkness of the dot indicating the number of times it was tested). Individual biases cluster around zero (from –2.2% to 1.2%). For each participant, alpha (a) is the bias of the logistic function, and beta (b) is the slope.

The individual results are plotted in Fig.40C. Each blue dot represents a size change that was tested by the Quest+ maximum likelihood model, and the darkness of the dot indicates the number of times it was tested. We then fit with a four-parameter logistic Bayesian psychometric function to each individual set of data (indicated with a black line), using the Palamedes Toolbox 1.10.1 (Prins

& Kingdom, 2018) with CmdStan 2.22.0, using the toolbox's standard priors (bias and slope: normal (0,100), upper and lower lapse rates: beta (1,10)), and based on 15,000 posterior estimates (100 posterior estimates are illustrated in red). Fig.40C shows that individual biases ranging from –2.2% to +1.2%, but clustered around 0.

To estimate the population level psychometric function illustrated in Fig.40A, we used the Palamedes Toolbox 1.10.1 (Prins & Kingdom, 2018) and CmdStan 2.22.0 to fit a four-parameter logistic hierarchical Bayesian psychometric function, which fits the data with a multilevel model that takes into account the variability of each subject. We used the toolbox's standard multilevel priors which are documented by Prins & Kingdom (2019) and, based on 15,000 posterior estimates (100 posterior estimates are illustrated in red), found a population level bias of –0.219% (95% CI: –1.82% to 1.39%) and a population level slope of –0.732 (95%CI: –1.07 to 0.378).

The estimate that particularly interests us is the population bias, so in Fig.40B we provide a probability density function of the 15,000 posterior estimates of the bias. We found no statistically significant bias, and therefore no statistically significant effect of vergence on perceived size. Indeed, the non-significant bias of –0.2% is in the wrong direction for size constancy.

To go beyond the negative claim that we found no statistically significant effect (null hypothesis not rejected) to the positive claim that there is no effect of vergence on perceived size (null hypothesis accepted), we can make two arguments.

First, from a Bayesian perspective, we can perform a JZS Bayes factor (Rouder et al., 2009). The estimated Bayes factor that we found was 3.99 (±0.03%), which suggests that the data are four times more likely under the null hypothesis (bias = 0) than under the alternative (bias ≠ 0).

Second, from a frequentist perspective, we can perform an inferiority test that tests whether any true vergence size constancy effect is at least as large as the smallest effect size of interest (Lakens et al., 2018). You'll remember, we define our smallest effect size of interest as the detection threshold for our most sensitive observer (which is 1.43%). Put simply, any vergence size constancy effect that's smaller than a 1.43% size change won't be detected by any of our observers. Since we have a directional hypothesis (vergence micropsia should reduce, rather than increase, the apparent size of the target), we specifically test whether there is a bias > 1.43%. We therefore perform an inferiority test by taking the 90% confidence interval of the population bias in Fig.40B in the predicted direction. This is 0.96%, and since it is smaller than 1.43% (our smallest effect size of interest), from a frequentist perspective we can conclude that any vergence size constancy effect is effectively equivalent to zero (Lakens et al., 2018).

Discussion

According to the literature, "it is well known that vergence is a reliable source of depth information for size constancy" (Sperandio et al., 2013), and yet our findings contradict this conclusion. We find no evidence that vergence makes any contribution to size constancy. To our knowledge, ours is the first study to report a failure of vergence size constancy at near distances. But to our knowledge, ours is also the first study that specifically controls for confounding perceptual cues (changes in the retinal image), whilst also controlling for confounding cognitive cues (keeping subjects naïve about changes in absolute distance). So this study should lead us to question vergence size constancy. This work has three important implications:

1. Visual Scale: First, it helps to challenge our understanding of visual scale. The scale of the visual scene is meant to be provided by a number of well documented distance cues, such as vergence, accommodation, motion parallax (how the scene moves with the motion of the observer), the familiar size of objects in the scene, and the location of the object relative to the ground-plane. But both accommodation (Mon-Williams & Tresilian, 2000) and motion parallax (Renner et al., 2013) have been found to be largely ineffectual as absolute distance cues, familiarity is merely thought to affect our cognitive judgement of size, rather than our actual perception (Gogel, 1969; Predebon, 1992), and the ground-plane only applies to limited viewing conditions and far distances (Creem-Regehr et al., 2015). Given these shortcomings, vergence was meant to

provide a solid anchor for our size and distance judgements in near space, that is necessary for effective reaching and grasping (Culham et al., 2008). But our results challenge this conclusion, leaving us in search for a new approach to visual scale.

2. Neural Processing: Second, our results help us to resolve the question posed by Chen et al. (2019) on the neural processing of size constancy. Ever since Trotter et al. (1992) found that a large majority of neurons in the primary visual context (V1) were modulated by the viewing distance provided by vergence, it has been suggested that processing of the vergence signal in the visual cortex plays an important role in size constancy (Trotter et al., 1993): "The finding that visual responsiveness is modulated by the viewing distance as early as the primary visual cortex indicates that integration of information from both retinal and extraretinal sources can occur early in the visual processing pathway for cortical representation of three-dimensional space." There is no question size constancy occurs in V1 (Murray et al., 2006; Sperandio et al., 2012), however Chen et al. (2019) recently found that size constancy takes ~150ms to evolve, suggesting that the vergence signal and the retinal image are not integrated during initial processing in V1 (~50ms). Instead, Chen et al. (2019) suggest that their findings are consistent with either (a) an integration of the vergence signal and the retinal image during subsequent processing in V1, and/or (b) top-down processing from higher-level visual areas to V1. Given our finding that vergence has no effect on perceived size, an account of size constancy based on the integration of the vergence distance with the retinal image no longer seems sustainable, and our results suggest that size constancy is likely to be much more reliant on top-down processes than previously thought.

3. Multisensory Integration: The Taylor illusion, where an after-image of the hand in darkness appears to shrink or grow with physical hand movements, has been an important topic for recent discussions of multisensory integration (Faivre et al., 2017a; Grove et al., 2019). According to Sperandio et al. (2013), the brain integrates the retinal image with proprioceptive information from vergence (major signal) and proprioceptive information from the hand (minor signal), so that when vergence and the hand move in the same direction we experience almost

perfect size constancy, and when they move in opposite directions we experience only a slight reduction in size constancy from vergence. But in light of our results, this explanation is no longer sustainable. Vergence doesn't explain size constancy in the Taylor illusion. But nor can hand movements, since when vergence and the hand move in opposite directions, it is vergence that dominates size constancy.

Instead, a new, purely cognitive, explanation for the Taylor illusion begins to emerge. In both conditions in Sperandio et al. (2013), subjects have conscious knowledge about their changing gaze position. When vergence and the hand move in the same direction, subjects' knowledge about their changing gaze position comes from their hand. When vergence and the hand move in opposite directions, subjects' knowledge about their changing gaze position comes from the motion in depth of the LED through a combination of retinal slip and/or proprioceptive sensation. And this alternative cognitive explanation suggests that rather than the Taylor illusion being due to a change in their visual experience (the perceived angular size of the after-image), subjects merely cognitively attribute this change in distance, and therefore a change in physical size, to a constant visual experience (an after-image with a constantly perceived angular size).

To give an analogy, consider how *visual* (optic flow) and *kinesthetic* (vestibular) cues interact in the experience of illusory self-motion. Vestibular cues are not necessary for illusory self-motion (Fischer & Kornmüller, 1930), e.g. when we see a neighbouring train pull away in a railway station whilst we are stationary. Nonetheless, the addition of vestibular acceleration can contribute to the illusion of self-motion (Ash et al., 2011). But, and this is the crucial point, there is no suggestion that this contribution has to be *visual*. It is not as if the optic flow has to be seen as flowing faster in order to incorporate this vestibular cue into our determination of self-motion. Instead, the visual (optic flow) and the non-visual (vestibular) cues merely feed into a common post-perceptual determination. Indeed, as Ash, Palmisano, & Kim (2011) observe, there is no 'mandatory fusion' between the *visual* (optic flow) and *kinesthetic* (vestibular) cues, which is often taken to be the litmus test of a truly perceptual effect: see Hillis, Ernst, Banks, & Landy (2002).

What would challenge this cognitive explanation is if the Taylor illusion persisted even in the absence of conscious knowledge about distance changes. But this is exactly what we tested in the context of changes in gaze position (which, after all, was the main signal for the Taylor illusion in Sperandio et al., 2013) and found no effect. So just because subjects use their conscious knowledge about their changing gaze position to attribute physical distance, and therefore physical size, to the afterimage, doesn't mean that that the visual system itself "relies on multimodal signals" (Sperandio et al., 2013; Chen et al., 2018) or that "visual consciousness is shaped by the body" (Faivre et al., 2015; Faivre et al., 2017a; Faivre et al., 2017b) (see Linton, 2017, pp.37-38 and pp.65-66 for a related discussion of multimodal 3D shape perception). This cognitive approach also suggests an alternative explanation for the integration of vision with the rubber-hand illusion (Faivre et al., 2017a) and tool use (Grove et al., 2019) in variants of the Taylor illusion.

9. The Paradox of Visual Scale

Our final discussion draws together three points. First, in our three experiments we demonstrate that the visual system is unable to extract absolute distance information from the environment using vergence (the angular rotation of the eyes), so this is not the means by which we scale the visual scene. Second, we have also considered the literature demonstrating that accommodation, motion parallax, and vertical disparities are largely ineffective as absolute distance cues, and that the ground plane and visual field blur are limited in application. Third, we have also considered the argument made by Walter Gogel and John Predebon that as a cue to absolute size and distance familiar size is merely cognitive in nature, and does not directly affect our perception. Coupled with the fact that little work on vista space perception effectively differentiates between the ground plane and familiar size, at it opens up the suggestion that what is determining visual scale isn't the triangulation of absolute distance by the visual system, but instead merely the cognitive processing of pictorial cues such as familiar size.

This appears to be the right conclusion to come to, but we face the following dilemma. So far we have only admitted the effectiveness (in a cognitive sense, at least) of Pictorial Cues to scale: 1. Familiar Size, 2. Ground Plane, and 3. Visual Field Blur. And rejected the effectiveness of Triangulation Cues to scale: 1. Vergence, 2. Accommodation, 3. Motion Parallax, 4. Vertical Disparities. (Although some discussions describe visual field blur as also being a triangulation cue, as Vishwanath & Blaser, 2010 observe, participants tend to double count visual field blur as making the surface both closer and more rotated, which is inconsistent with the triangulation account). However, my account runs into the following dilemma. We know that triangulation cues dominate pictorial cues to scale. As you increase the IPD in binocular viewing of the real world, you shrink the apparent scale of the visual scene. This is something that Helmholtz (1858) articulated in the context of Telestereoscopic viewing where the IPD is increased:

> "…it seems as if the observer were looking not at the natural landscape itself, but at a very exquisite and exact model of it, reduced in scale."

You can experience the effect for yourself looking at hyper-stereoscopic photographs, and a couple of examples are provided in Figure 41 and Figure 42. But what explains this effect?

Clearly it is not pictorial cues, as they remain unaffected, as closing one eye will confirm. Nor does it appear to be motion parallax. Instead, explanations have focused on vergence (Helmholtz, 1858; 1866) or a combination of vergence and vertical disparities (Rogers, 2011). But I have argued that neither vergence nor vertical disparities should be thought of as effective absolute distance cues, hence the paradox.

[Removed due to copyright]

Figure 41. Twin Stop 3D (2009) by Sasha Becher. © Sasha Becher. https://www.fotocommunity.com/photo/twin-stop-3d-anaglyph-sasha-becher/16025031 For more images see https://www.fotocommunity.com/user_photos/789955

[Removed due to copyright]

Figure 42. Traffic on the Autobahn (2011) by Sasha Becher. © Sasha Becher. https://www.flickr.com/photos/stereotron/6603863175/in/album-72157673910516235/ For more images see https://www.flickr.com/photos/stereotron/

So the question is what provides the additional information? In Linton (2018), I argue by a process of elimination. In Telestereoscopic viewing we manipulate three things: (1) vergence, (2) vertical disparities, and (3) horizontal disparities. Since the change in apparent scale cannot be due to (1) or (2), it must be due to (3). But how can horizontal disparities be responsible?

Traditionally, horizontal disparities are thought of as merely relative depth cues. Indeed, even this claims too much. Angular size is a relative depth cue, since angular size varies in proportion to distance. But since disparity varies in proportion to distance^2, even once angular size is controlled for, (a) disparity, is like (b) accommodation, (c) defocus blur, and (d) vergence, varying with 1/distance (disparity/vergence are the retinal/extra-retinal manifestations of the same property, in much the same way that accommodation/defocus blur are). As Brenner & van Damme (1998) note about vergence, a 1° change could be due to a shift from 20cm to 21cm, or 2m to 4m. The same is true of binocular disparity. Does 1° of disparity signify a doubling of distance (2m to 4m), or merely an increase by 1/20 (20cm to 21cm)? Without knowing the initial distance of the fixation plane (typically suggested via vergence), we have no way of knowing.

My solution to the paradox is to invert the traditional relationship between depth constancy from binocular disparity and 3D shape perception. Typically the focus in the literature is on adding increasing numbers of depth cues in the attempt of to show that something close to depth constancy (retaining the 3D shape of an object with distance) can be achieved from binocular disparity (Guan & Banks, 2016). This requires positing sources of absolute distance information by which the visual system can scale the z-axis depth from disparity (that reduces in proportion to the distance squared) relative to x-axis and y-axis angular size (which reduces in proportion to the distance).

My argument is that we should invert this way of thinking about the problem. Instead, we should accept that there is no depth constancy; that perceived 3D shape from disparity really does fall off with distance. However, we can use this failure of depth constancy, i.e. fall off of 3D shape from disparity, to scale the distance of the object so long as we know its 3D shape. So instead of going from distance → 3D shape, we go from 3D shape → distance.

This is where the distinction between triangulation and pictorial cues re-emerges. The 3D shape of objects from perspective is distorted slightly with distance, especially at closer distances. The classic example is the relationship between focal length and the perceived shape of the face in portraiture (Fig.43), but is also exploited by the dolly zoom effect pioneered by Irmin Roberts for Alfred Hichcock's film Vertigo (1958), where the geometry of the scene itself, and not just an object in the scene, is distorted by the change in viewing distance.

[Removed due to copyright]

Figure 43. Example of how viewing distance affects the perceived shape from perspective. Although this effect is often attributed to a change in focal length of the lens, it is the fact that the viewing distance has changed that causes this effect, rather than being an artefact of the change in the focal length of the lens. Source: https://www.diyphotography.net/gif-explains-changing-focal-length-impacts-portrait/

However, overall these perspective distortions from changes in distance are minor in comparison to the drastic fall off in disparity with the square of the distance. This means that if we have an idea of the 3D shape of the object from perspective or shading, we can compare it with the 3D shape from binocular disparity, to get an impression of the distance of the object or scene.

This would have to rely on perceptual learning. Over time we learn to develop expectations of the kind of binocular stereopsis we expect to experience from objects and scenes at different viewing distances. This could be quite specific. For instance, by regularly interacting with familiar objects and familiar scene geometries (like corridors and rooms), we develop a very heavily context dependent appreciation of binocular scale. Or, in the context of unfamiliar scenes, it could be a looser relationship and more reliant on natural scene statistics; the notion that we only experience certain kinds of binocular stereopsis at certain scales.

I developed this proposal in Linton (2018). Independently, Brenner & Smeets (2018) also raise the possibility that if we know the 3D shape of an object we can calculate its distance from the fall off of its binocular disparity with distance. But that is where the similarity between our positions ends. My argument has four stages:

1. We can in theory extract distance from the fall off of binocular stereopsis if we already know the geometry of the object (this is in common with Brenner & Smeets, 2018)

2. We can do this in one of three ways:
    a. Reliance on knowledge about the shape of 3D objects in the scene (this is in common with Brenner & Smeets, 2018)
    b. Reliance on knowledge about the geometry of the scene itself (the 3D shape of the scene itself, and the relative depth of objects within the scene)
    c. Reliance on natural scene statistics (knowledge about typically which kinds of binocular stereopsis we experience in which scale contexts)

3. This mechanism is what underpins binocular scale perception. This is only an argument that can be made in light of my experimental work on vergence. Brenner & Smeets (2018) regard this as one of many potential absolute distance cues.

4. This is <u>not</u> the perception of scale, but merely its cognition. The point is this. So far as our visual experience is concerned, all that we experience visually when we reduce the amount of binocular disparity within a scene is an impression that the scene is flatter. All we see is a change in the 3D geometry of the scene (accentuation or flattening), which we cognitively associate with a change in scale.

We shouldn't make the mistake of claiming that this is the exclusive basis of visual scale. People who don't have binocular coordination are still able to navigate through the world without any major impediment (such as Sue Barry (SB): see Sacks, 2006 and Barry, 2009); with many of them not noticing that they have any visual deficits until they are diagnosed much later in life; although those without functioning binocular coordination do have compromised reach and grasping motions at near distances (see Melmoth & Grant, 2006). Instead, we need to recognise that binocular vision affords us a new way of gauging visual scale that supersedes the monocular cues that e.g. Sue Barry relied upon. But there is no difference in kind. We wouldn't want to say (and shouldn't say) that monocular scale is merely cognitive, but binocular scale is perceptual.

The monocular cues, such as familiarity, that Sue Barry relied upon are the cues to scale that we rely upon when we watch a movie or play a computer game. We have no difficulty with the size and distance of objects when watching movies and television, and this suggests that, in order for movies and television to have been successful in the first place, there must be some similarity between pictorial scale (films / television / computer games) and visual scale. Indeed, for monocular viewing there seems to be no distinction between pictorial and visual scale. 'Seeing' the scale of the statue of liberty in person is no different from 'seeing' the scale of the statue of liberty in a photo. And note how rarely an extended ground-plane is a component of the movies and television shows whose scale we instantly recognise without difficulty.

Binocular vision could have provided the visual system with a range finder by means of vergence or vertical disparities. But the results of my work on vergence, when coupled with the failure to find that vertical disparities are effective absolute distance cues when properly tested (Cumming et al., 1991; Sobel & Collett, 1991), with the only evidence in their favour coming from when they are improperly tested (Rogers & Bradshaw, 1995), leaves us with the conclusion that there is, in fact, no such binocular range finding capabilities.

Finally, in order to explain the binocular contribution to the 'perception' of scale I have proposed that observers are responsive to the way in which the 3D shape from binocular stereopsis varies with distance when compared with 3D shape from pictorial cues (perspective, shading, familiarity) which are largely invariant to changes with distance (perspective, too, changes if the relative distance between the observer and the front and the back of the object change significantly; although not as much as the change in disparity which is a function of distance^2).

Conclusions

The argument of this thesis is that the visual system does not appear to have access to absolute distance information, and we should therefore be open to the idea that vision does not directly provide us with absolute scale. This argument rests on the following propositions:

1. The effectiveness of motion parallax is widely questioned in the literature [Ch.2]

2. Evidence supporting vertical disparities only applies to very limited viewing conditions, and arguably relies on misrepresenting them [Ch.2]

3. Evidence supporting vergence appears to rely on introducing confounding cues [Ch.2]

4. Vergence is not an effective absolute distance cue once these confounding cues have been controlled for [Ch.4]

5. This ineffectiveness cannot be attributed to vergence / accommodation conflict [Ch.5]

6. Vergence is not an effective cue to size constancy [Ch.8]

7. Interim Conclusion: This suggests that the visual system primarily relies on 'cognitive' cues to absolute distance such as familiar size [Ch.9]. The ground plane may also play a role, but (a) there has been little work to distinguish the ground plane from familiar size, and, when there is, (b) there is little to show that the spatial biases that emerge are anything other than cognitive biases.

8. The problem with this conclusion is the 'paradox of visual scale': Triangulation cues appear to be ineffective, and yet we know from Telestereoscopic viewing (where increasing the IPD appears to shrink the visual scene in binocular viewing) that triangulation cues dominate pictorial cues such as familiar size [Ch.9]

9. The solution is to suggest that binocular scale is gauged by comparing the perceived 3D shape from binocular disparity with the perceived shape either (a) from monocular cues, or (b) from natural scene statistics [Ch.9]. To give an example, an observer viewing a cylinder will be able to recognise its circular cross section from a combination of pictorial cues (perspective, shading) and experience (as well as assumptions, that if it looks roughly circular, it is), and automatically contrast with the perception of its depth from binocular disparity, with accentuated depth at near distances, and reduced depth at far distances, as found by Johnston (1991).

References

Abbott, A. L., & Ahuja, N. (1990). Active surface reconstruction by integrating focus, vergence, stereo, and camera calibration. *[1990] Proceedings Third International Conference on Computer Vision*, 489–492. https://doi.org/10.1109/ICCV.1990.139576

Abbott, A. L., & Ahuja, N. (1988). Surface Reconstruction By Dynamic Integration Of Focus, Camera Vergence, And Stereo. *[1988 Proceedings] Second International Conference on Computer Vision*, 532–543. https://doi.org/10.1109/CCV.1988.590034

Adams, O. S. (1955). Stereogram decentration and stereo-base as factors influencing the apparent size of stereoscopic pictures. *The American Journal of Psychology*, *68*(1), 54–68.

al-Haytham, Ibn. (1021). *Kitab Al-Manazir (Book of Optics). In A. I. Sabra (Ed.), The Optics of Ibn al-Haytham (1983). Kuwait: National Council for Culture, Arts and Letters.*

Ash, A., Palmisano, S., & Kim, J. (2011). Vection in Depth during Consistent and Inconsistent Multisensory Stimulation. *Perception*, *40*(2), 155–174. https://doi.org/10.1068/p6837

Auvray, M., & O'Regan, J. (2003). L'influence des facteurs sémantiques sur la cécité aux changements progressifs dans les scènes visuelles. *L'Année psychologique*, *103*(1), 9–32. https://doi.org/10.3406/psy.2003.29621

Baird, J. W. (1903). The Influence of Accommodation and Convergence upon the Perception of Depth. *The American Journal of Psychology*, *14*(2), 150–200. https://doi.org/10.2307/1412712

Banks, M. S., Hoffman, D. M., Kim, J., & Wetzstein, G. (2016). 3D Displays. *Annual Review of Vision Science*, *2*(1), 397–435. https://doi.org/10.1146/annurev-vision-082114-035800

Barry, S. R. (2009). *Fixing My Gaze: A Scientist's Journey Into Seeing in Three Dimensions*. Hachette UK.

Bates, D. M., Maechler, M., & Bolker, B. (2012). *Lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.*

Beall, A. C., Loomis, J. M., Philbeck, J. W., & Fikes, T. G. (1995). Absolute motion parallax weakly determines visual scale in real and virtual environments. *Human Vision, Visual Processing, and Digital Display VI*, *2411*, 288–297. https://doi.org/10.1117/12.207547

Berkeley, G. (1709). *An Essay Towards a New Theory of Vision*. Printed by Aaron Rhames, at the Back of Dick's Coffee-House, for Jeremy Pepyat, Bookseller in Skinner-Row. https://www.maths.tcd.ie/~dwilkins/Berkeley/Vision/1709A/Vision.pdf

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(4), 1076–1096. https://doi.org/10.1037//0096-1523.25.4.1076

Biersdorf, W. R., Ohwaki, S., & Kozil, D. J. (1963). The Effect of Instructions and Oculomotor Adjustments on Apparent Size. *The American Journal of Psychology*, *76*(1), 1–17. JSTOR. https://doi.org/10.2307/1419994

Bishop, P. O. (1989). Vertical Disparity, Egocentric Distance and Stereoscopic Depth Constancy: A New Interpretation. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *237*(1289), 445–469.

Bishop, P. O., & Pettigrew, J. D. (1986). Neural mechanisms of binocular vision. *Vision Research*, *26*(9), 1587–1600. https://doi.org/10.1016/0042-6989(86)90177-X

Bock, O. (1986). Contribution of retinal versus extraretinal signals towards visual localization in goal-directed movements. *Experimental Brain Research*, *64*(3), 476–482. https://doi.org/10.1007/bf00340484

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York;London: D. Appleton-Century Co. https://trove.nla.gov.au/version/44237362

Bradshaw, M. F., Elliott, K. M., Watt, S. J., Hibbard, P. B., Davies, I. R. L., & Simpson, P. J. (2004). Binocular cues and the control of prehension. *Spatial Vision*, *17*(1–2), 95–110.

Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research*, *40*(27), 3725–3734. https://doi.org/10.1016/S0042-6989(00)00214-5

Brainard, D. H. (2017). *mQUESTPlus: A Matlab implementation of QUEST+*. https://github.com/brainardlab/mQUESTPlus

Brenner, E., & Smeets, J. B. (2018). Depth Perception. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience: Sensation, Perception, and Attention*, 385–414. https://doi.org/10.1002/9781119170174

Brenner, E., & van Damme, W. J. (1998). Judging distance from ocular convergence. *Vision Research*, *38*(4), 493–498.

Brenner, E., Van Den Berg, A. V., & Van Damme, W. J. (1996). Perceived motion in depth. *Vision Research*, *36*(5), 699–706.

Brewster, D. (1844). On the knowledge of distance given by binocular vision. *Transactions of the Royal Society of Edinburgh*, *15*, 663–674.

Bross, M. (2000). Emmert’s law in the dark: Active and passive proprioceptive effects on positive visual afterimages. *Perception*, *29*(11), 1385–1391. https://doi.org/10.1068/p3038

Brown, C., Coombs, D., & Soong, J. (1992). Real-time Smooth Pursuit Tracking. In A. Blake & A. L. Yuille, *Active Vision* (pp. 123–136). MIT Press.

Campagnoli, C., Croom, S., & Domini, F. (2017). Stereovision for action reflects our perceptual experience of distance and depth. *Journal of Vision*, *17*(9), 21–21. https://doi.org/10.1167/17.9.21

Campagnoli, C., & Domini, F. (2019). Does depth-cue combination yield identical biases in perception and grasping? *Journal of Experimental Psychology: Human Perception and Performance*, *45*(5), 659–680. https://doi.org/10.1037/xhp0000636

Carey, D. P., & Allan, K. (1996). A motor signal and “visual” size perception. *Experimental Brain Research*, *110*(3), 482–486. https://doi.org/10.1007/BF00229148

Charman, W. N., & Heron, G. (2015). Microfluctuations in accommodation: An update on their characteristics and possible role. *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists)*, *35*(5), 476–499. https://doi.org/10.1111/opo.12234

Chen, J., Sperandio, I., & Goodale, M. A. (2018). Proprioceptive Distance Cues Restore Perfect Size Constancy in Grasping, but Not Perception, When Vision Is Limited. *Current Biology: CB*, *28*(6), 927-932.e4. https://doi.org/10.1016/j.cub.2018.01.076

Chen, J., Sperandio, I., Henry, M. J., & Goodale, M. A. (2019). Changing the Real Viewing Distance Reveals the Temporal Evolution of Size Constancy in Visual Cortex. *Current Biology*, *29*(13), 2237-2243.e4. https://doi.org/10.1016/j.cub.2019.05.069

Choe, K. W., Blake, R., & Lee, S.-H. (2016). Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vision Research*, *118*, 48–59. https://doi.org/10.1016/j.visres.2014.12.018

Cholewiak, S. A., Love, G. D., & Banks, M. S. (2018). Creating correct blur and its effect on accommodation. *Journal of Vision*, *18*(9), 1–1. https://doi.org/10.1167/18.9.1

Cholewiak, S. A., Love, G. D., Srinivasan, P. P., Ng, R., & Banks, M. S. (2017). Chromablur: Rendering chromatic eye aberration improves accommodation and realism. *ACM Transactions on Graphics*, *36*(6), 210:1–210:12. https://doi.org/10.1145/3130800.3130815

Clark, J. J., & Yuille, A. L. (1990). *Data Fusion for Sensory Information Processing Systems*. Springer.

Clément, G., Loureiro, N., Sousa, D., & Zandvliet, A. (2016). Perception of Egocentric Distance during Gravitational Changes in Parabolic Flight. *PLOS ONE*, *11*(7), e0159422. https://doi.org/10.1371/journal.pone.0159422

Collewijn, H., & Erkelens, C. J. (1990). Binocular eye movements and perception of depth. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes* (pp. 213–261). Elsevier.

Combe, E., & Wexler, M. (2010). Observer Movement and Size Constancy. *Psychological Science*, *21*(5), 667–675. JSTOR.

Coombs, D., & Brown, C. (1992). Real-time smooth pursuit tracking for a moving binocular robot. *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 23–28. https://doi.org/10.1109/CVPR.1992.223232

Coombs, David, & Brown, C. (1993). Real-time binocular smooth pursuit. *International Journal of Computer Vision*, *11*(2), 147–164. https://doi.org/10.1007/BF01469226

Creem-Regehr, S. H., Stefanucci, J. K., & Thompson, W. B. (2015). Chapter Six - Perceiving Absolute Scale in Virtual Environments: How Theory and Application Have Mutually Informed the Role of Body-Based Perception. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 62, pp. 195–224). Academic Press. https://doi.org/10.1016/bs.plm.2014.09.006

Culham, J., Gallivan, J., Cavina-Pratesi, C., & Quinlan, D. (2008). FMRI investigations of reaching and ego space in human superior parieto-occipital cortex. In *Embodiment, Ego-space and Action* (pp. 247–274).

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1991). Vertical Disparities and Perception of Three-Dimensional Shape. *Nature; London*, *349*(6308), 411–413.

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). Academic Press. https://doi.org/10.1016/B978-012240530-3/50005-5

Darwin, R. W. (1786). XVI. New experiments on the ocular spectra of light and colours. *Philosophical Transactions of the Royal Society of London*, *76*, 313–348. https://doi.org/10.1098/rstl.1786.0016

de la Malla, C., Buiteman, S., Otters, W., Smeets, J. B. J., & Brenner, E. (2016). How various aspects of motion parallax influence distance judgments, even when we think we are standing still. *Journal of Vision*, *16*(9), 8. https://doi.org/10.1167/16.9.8

Descartes, R. (1637). *Dioptrique (Optics). In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), The Philosophical Writings of Descartes: Volume 1 (1985)*. Cambridge University Press.

Dobbins, A. C., Jeo, R. M., Fiser, J., & Allman, J. M. (1998). Distance Modulation of Neural Activity in the Visual Cortex. *Science*, *281*(5376), 552–555. https://doi.org/10.1126/science.281.5376.552

Drewes, J., Zhu, W., Hu, Y., & Hu, X. (2014). Smaller Is Better: Drift in Gaze Measurements due to Pupil Dynamics. *PLOS ONE*, *9*(10), e111197. https://doi.org/10.1371/journal.pone.0111197

Durgin, F. H. (2017). Counterpoint: Distinguishing between perception and judgment of spatial layout. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *12*(2), 344–346. https://doi.org/10.1177/1745691616677829

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review*, *16*(5), 964–969. https://doi.org/10.3758/PBR.16.5.964

Emmert, E. (1881). Größenverhältnisse der Nachbilder. *Klinische Monatsblätter Für Augenheilkunde Und Für Augenärztliche Fortbildung*, *19*, 443–450.

Enright, J. T. (1989). Enright JT (1989). "The eye, the brain and the size of the moon: Toward a unified oculomotor hypothesis for the moon illusion". In Hershenson, Maurice (ed.). The Moon illusion. Hillsdale, N.J: L. Erlbaum Associates. Pp. 59–121. In M. Hershenson, *The Moon Illusion* (pp. 59–121). L. Erlbaum Associates.

Erkelens, C. J., & Collewijn, H. (1985b). Eye movements and stereopsis during dichoptic viewing of moving random-dot stereograms. *Vision Research*, *25*(11), 1689–1700. https://doi.org/10.1016/0042-6989(85)90141-5

Erkelens, C. J., & Collewijn, H. (1985a). Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Research*, *25*(4), 583–588.

Faivre, N., Arzi, A., Lunghi, C., & Salomon, R. (2017b). Consciousness is more than meets the eye: A call for a multisensory study of subjective experience. *Neuroscience of Consciousness*, *2017*(1). https://doi.org/10.1093/nc/nix003

Faivre, N., Dönz, J., Scandola, M., Dhanis, H., Ruiz, J. B., Bernasconi, F., Salomon, R., & Blanke, O. (2017a). Self-Grounded Vision: Hand Ownership Modulates Visual Location through Cortical β and γ Oscillations. *Journal of Neuroscience*, *37*(1), 11–22. https://doi.org/10.1523/JNEUROSCI.0563-16.2016

Faivre, N., Salomon, R., & Blanke, O. (2015). Visual consciousness and bodily self-consciousness. *Current Opinion in Neurology*, *28*(1), 23–28. https://doi.org/10.1097/WCO.0000000000000160

Firestone, C. (2013). How "Paternalistic" Is Spatial Perception? Why Wearing a Heavy Backpack Doesn't—and Couldn't—Make Hills Look Steeper: *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691613489835

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *The Behavioral and Brain Sciences*, *39*, e229. https://doi.org/10.1017/S0140525X15000965

Firestone, C., & Scholl, B. J. (2017). Seeing and Thinking in Studies of Embodied "Perception." *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(2), 341–343. https://doi.org/10.1177/1745691616679944

Fischer, M. H., & Kornmüller, A. E. (1930). Optokinetic-Induced Motion Perception and Optokinetic Nystagmus. *Journal of Psychology and Neurology (Leipzig)*, *41*, 273–308.

Fisher, S. K., & Ciuffreda, K. J. (1988). Accommodation and apparent distance. *Perception*, *17*(5), 609–621. https://doi.org/10.1068/p170609

Foley, J. (1980). Binocular Distance Perception. *Psychological Review*, *87*, 411–434. https://doi.org/10.1037//0033-295X.87.5.411

Foley, J. M., Ribeiro-Filho, N. P., & Da Silva, J. A. (2004). Visual perception of extent and the geometry of visual space. *Vision Research*, *44*(2), 147–156. https://doi.org/10.1016/j.visres.2003.09.004

Frank, H. (1930). Ueber den Einfluss inadaquater Konvergenz und Akkommodation auf die Sehgrosse. *Psychol. Forsch.*, *13*, 135–144.

Fried, A. H. (1973). *Convergence as a cue to distance* [PhD Dissertation]. New School of Social Research.

Frost, B. J. (2009). Bird head stabilization. *Current Biology*, *19*(8), R315–R316. https://doi.org/10.1016/j.cub.2009.02.002

Gallivan, J. P., Cavina-Pratesi, C., & Culham, J. C. (2009). Is That within Reach? FMRI Reveals That the Human Superior Parieto-Occipital Cortex Encodes Objects Reachable by the Hand. *Journal of Neuroscience*, *29*(14), 4381–4391. https://doi.org/10.1523/JNEUROSCI.0377-09.2009

Geiger, D., & Yuille, A. (1989). Stereo and Eye Movement. *Biol. Cybern.*, *62*(2), 117–128. https://doi.org/10.1007/BF00203000

Gerd Neumann Jr. (n.d.). *Gerd Neumann: Deep black optical paint*. Retrieved June 21, 2018, from https://www.gerdneumann.net/english/instrument-building-parts-teile-fuer-den-fernrohrbau/totmatte-schwarze-optikfarbe-deep-black-optical-paint.html

Gibson, J. J. (1947). *Motion Picture Testing and Research*. Army Air Forces Washington DC Aviation Psychology Program. http://www.dtic.mil/docs/citations/AD0651783

Gibson, J. J. (1950). *The Perception of the Visual World*. Houghton Mifflin Co.

Gillam, B., Chambers, D., & Lawergren, B. (1988). The role of vertical disparity in the scaling of stereoscopic depth perception: An empirical and theoretical study. *Perception & Psychophysics*, *44*(5), 473–483. https://doi.org/10.3758/BF03210433

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic Depth Constancy Depends on the Subject’s Task. *Vision Research*, *36*(21), 3441–3456. https://doi.org/10.1016/0042-6989(96)00090-9

Glennerster, Andrew, Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology: CB*, *16*(4), 428–432. https://doi.org/10.1016/j.cub.2006.01.019

Glover, S., Wall, M. B., & Smith, A. T. (2012). Distinct cortical networks support the planning and online control of reaching-to-grasp in humans. *European Journal of Neuroscience*, *35*(6), 909–915. https://doi.org/10.1111/j.1460-9568.2012.08018.x

Gnadt, J. W., & Mays, L. E. (1995). Neurons in monkey parietal area LIP are tuned for eye-movement parameters in three-dimensional space. *Journal of Neurophysiology*, *73*(1), 280–297. https://doi.org/10.1152/jn.1995.73.1.280

Goethe, J. W. von. (1810). *Zur Farbenlehre*. in der J.G. Cotta’schen Buchhandlung.

Gogel, W. C. (1962). The Effect of Convergence on Perceived Size and Distance. *The Journal of Psychology*, *53*(2), 475–489. https://doi.org/10.1080/00223980.1962.9916591

Gogel, W. C. (1969). The Sensing of Retinal Size. *Vision Research*, *9*, 1079–1094.

Gogel, W. C. (1976). An indirect method of measuring perceived distance from familiar size. *Perception & Psychophysics*, *20*(6), 419–429. https://doi.org/10.3758/BF03208276

Gogel, W. C. (1998). An analysis of perceptions from changes in optical size. *Perception & Psychophysics*, *60*(5), 805–820. https://doi.org/10.3758/BF03206064

Gogel, W. C., & Da Silva, J. A. (1987). Familiar size and the theory of off-sized perceptions. *Perception & Psychophysics*, *41*(4), 318–328. https://doi.org/10.3758/BF03208233

Gogel, W. C., & Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception & Psychophysics*, *13*(2), 284–292. https://doi.org/10.3758/BF03214141

Goldstein, E. B., & Brockmole, J. (2016). *Sensation and Perception*. Cengage Learning.

Grant, S., & Conway, M. L. (2019). Some binocular advantages for planning reach, but not grasp, components of prehension. *Experimental Brain Research*, *237*(5), 1239–1255. https://doi.org/10.1007/s00221-019-05503-4

Gregory, R. L., & Ross, H. E. (1964). Visual Constancy during Movement: 1. Effects of S’s Forward and Backward Movement on Size Constancy. *Perceptual and Motor Skills*, *18*(1), 3–8. https://doi.org/10.2466/pms.1964.18.1.3

Gregory, R. L., Wallace, J. G., & Campbell, F. W. (1959). Changes in the size and shape of visual after-images observed in complete darkness during changes of position in space. *Quarterly Journal of Experimental Psychology*, *11*(1), 54–55. https://doi.org/10.1080/17470215908416288

Grove, C., Cardinali, L., & Sperandio, I. (2019). Does Tool-Use Modulate the Perceived Size of an Afterimage During the Taylor Illusion? [Abstract ECVP Leuven 2019]. *Perception*, *48*, 181.

Grove, P. M., Finlayson, N. J., & Ono, H. (2014). The Effect of Stimulus Size on Stereoscopic Fusion Limits and Response Criteria: *Perception*. https://doi.org/10.1068/p7513

Guan, P., & Banks, M. S. (2016). Stereoscopic depth constancy. *Phil. Trans. R. Soc. B*, *371*(1697), 20150253. https://doi.org/10.1098/rstb.2015.0253

Harris, J. M. (2006). The interaction of eye movements and retinal signals during the perception of 3-D motion direction. *Journal of Vision*, *6*(8), 2–2. https://doi.org/10.1167/6.8.2

Hatfield, G. (2002). Perception as Unconscious Inference. In D. Heyer & R. Mausfeld (Eds.), *Perception and the Physical World: Psychological and Philosophical Issues in Perception* (pp. 113–143). John Wiley and Sons.

Heinemann, E. G., Tulving, E., & Nachmias, J. (1959). The effect of oculomotor adjustments on apparent size. *The American Journal of Psychology*, *72*, 32–45. https://doi.org/10.2307/1420209

Held, R. T., Cooper, E. A., O’Brien, J. F., & Banks, M. S. (2010). Using Blur to Affect Perceived Distance and Size. *ACM Transactions on Graphics*, *29*(2). https://doi.org/10.1145/1731047.1731057

Helmholtz, H. (1857). das Telestereoskop. *Annalen Der Physik*, *178*(9), 167–175. https://doi.org/10.1002/andp.18571780907

Helmholtz, H. (1866). *Handbuch der Physiologischen Optik, Vol.III (translated by J. P. C. Southall 1925 Opt. Soc. Am. Section 26, reprinted New York: Dover, 1962).*

Helmholtz, P. H. (1858). II. On the telestereoscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *15*(97), 19–24. https://doi.org/10.1080/14786445808642433

Henriques, D. Y., & Crawford, J. D. (2000). Direction-dependent distortions of retinocentric space in the visuomotor transformation for pointing. *Experimental Brain Research*, *132*(2), 179–194. https://doi.org/10.1007/s002210000340

Henriques, D. Y. P., Medendorp, W. P., Gielen, C. C. a. M., & Crawford, J. D. (2003). Geometric computations underlying eye-hand coordination: Orientations of the two eyes and the head. *Experimental Brain Research*, *152*(1), 70–78. https://doi.org/10.1007/s00221-003-1523-4

Henriques, Denise Y. P., Klier, E. M., Smith, M. A., Lowy, D., & Crawford, J. D. (1998). Gaze-Centered Remapping of Remembered Visual Space in an Open-Loop Pointing Task. *Journal of Neuroscience*, *18*(4), 1583–1594. https://doi.org/10.1523/JNEUROSCI.18-04-01583.1998

Hermans, T. G. (1937). Visual size constancy as a function of convergence. *Journal of Experimental Psychology*, *21*(2), 145–161. https://doi.org/10.1037/h0058367

Hermans, Thomas G. (1954). The relationship of convergence and elevation changes to judgments of size. *Journal of Experimental Psychology*, *48*(3), 204–208. https://doi.org/10.1037/h0059355

Heuer, H. (1987). Apparent Motion in Depth Resulting from Changing Size and Changing Vergence. *Perception*, *16*(3), 337–350. https://doi.org/10.1068/p160337

Hillebrand, F. (1894). Das Verhältnis von Accommodation und Konvergenz zur Tiefenlokalisation. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, *7*, 97–151.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining Sensory Information: Mandatory Fusion Within, but Not Between, Senses. *Science*, *298*(5598), 1627–1630. https://doi.org/10.1126/science.1075396

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3), 33–33. https://doi.org/10.1167/8.3.33

Hooge, I. T. C., Hessels, R. S., & Nyström, M. (2019). Do pupil-based binocular video eye trackers reliably measure vergence? *Vision Research*, *156*, 1–9. https://doi.org/10.1016/j.visres.2019.01.004

Howard, I. P. (2008). Vergence modulation as a cue to movement in depth. *Spatial Vision*, *21*(6), 581–592. https://doi.org/10.1163/156856808786451417

Howard, I. P. (2012). *Perceiving in Depth, Volume 3: Other Mechanisms of Depth Perception*. Oxford University Press.

Howard, I. P., Fujii, Y., & Allison, R. S. (2014). Interactions between cues to visual motion in depth. *Journal of Vision*, *14*(2), 14–14. https://doi.org/10.1167/14.2.14

Jaschinski, W., Bröde, P., & Griefahn, B. (1999). Fixation disparity and nonius bias. *Vision Research*, *39*(3), 669–677. https://doi.org/10.1016/S0042-6989(98)00163-1

Jaschinski, W., Jainta, S., Hoormann, J., & Walper, N. (2007). Objective vs subjective measurements of dark vergence. *Ophthalmic and Physiological Optics*, *27*(1), 85–92. https://doi.org/10.1111/j.1475-1313.2006.00448.x

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*(7), 1351–1360. https://doi.org/10.1016/0042-6989(91)90056-B

Jones, J. A., Swan, J. E., II, Singh, G., & Ellis, S. R. (2011). Peripheral Visual Information and Its Effect on Distance Judgments in Virtual and Augmented Environments. *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, 29–36. https://doi.org/10.1145/2077451.2077457

Jones, J. A., Swan, J. E., II, Singh, G., Kolstad, E., & Ellis, S. R. (2008). The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception. *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, 9–14. https://doi.org/10.1145/1394281.1394283

Judd, C. H. (1897). Some facts of binocular vision. *Psychological Review*, *4*(4), 374–389. https://doi.org/10.1037/h0073701

Kepler, J. (1604). Paralipomena to Witelo. In W. H. Donahue (Trans.), *Optics: Paralipomena to Witelo and Optical Part of Astronomy.* Green Lion Press, 2000.

Kertesz, A. E., Hampton, D. R., & Sabrin, H. W. (1983). The unreliability of nonius line estimates of vertical fusional vergence performance. *Vision Research*, *23*(3), 295–297. https://doi.org/10.1016/0042-6989(83)90119-0

Kilpatrick, F. P., & Ittelson, W. H. (1953). The size-distance invariance hypothesis. *Psychological Review*, *60*(4), 223–231. https://doi.org/10.1037/h0060882

Kleiner, M., Brainard, D., & Pelli, D. (2007). Kleiner M, Brainard D, Pelli D, 2007, "What's new in Psychtoolbox-3?" Perception 36 ECVP Abstract Supplement. *Perception*, *36*(ECVP Abstract Supplement).

Klinghammer, M., Schütz, I., Blohm, G., & Fiehler, K. (2016). Allocentric information is used for memory-guided reaching in depth: A virtual reality study. *Vision Research*, *129*, 13–24. https://doi.org/10.1016/j.visres.2016.10.004

Koffka, K. (1935). *Principles of Gestalt psychology* (p. 720). Harcourt, Brace.

Kohly, R. P., & Ono, H. (2002). Fixating on the wallpaper illusion: A commentary on "The role of vergence in the perception of distance: A fair test of Bishop Berkeley's claim" by Logvinenko et al. (2001). *Spatial Vision*, *15*(3), 377–386.

Komoda, M. K., & Ono, H. (1974). Oculomotor adjustments and size-distance perception. *Perception & Psychophysics*, *15*(2), 353–360. https://doi.org/10.3758/BF03213958

Konrad, R., Angelopoulos, A., & Wetzstein, G. (2019). Gaze-Contingent Ocular Parallax Rendering for Virtual Reality. *ArXiv:1906.09740 [Cs]*. http://arxiv.org/abs/1906.09740

Krotkov, E., Fuma, F., & Summers, J. (1988). An agile stereo camera system for flexible image acquisition. *IEEE Journal on Robotics and Automation*, *4*(1), 108–113. https://doi.org/10.1109/56.782

Krotkov, E., Henriksen, K., & Kories, R. (1990). Stereo ranging with verging cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(12), 1200–1205. https://doi.org/10.1109/34.62610

Krotkov, E., & Kories, R. (1988). Adaptive control of cooperating sensors: Focus and stereo ranging with an agile camera system. *1988 IEEE International Conference on Robotics and Automation Proceedings*, 548–553 vol.1. https://doi.org/10.1109/ROBOT.1988.12109

Krotkov, E. P. (1989). *Active Computer Vision by Cooperative Focus and Stereo*. Springer-Verlag. https://doi.org/10.1007/978-1-4613-9663-5

Krotkov, Eric, & Bajcsy, R. (1993). Active vision for reliable ranging: Cooperating focus, stereo, and vergence. *International Journal of Computer Vision*, *11*(2), 187–203. https://doi.org/10.1007/BF01469228

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Lehky, S., Pouget, A., & Sejnowski, T. (1990). Neural Models of Binocular Depth Perception. *Cold Spring Harbor Symposia on Quantitative Biology*, *55*, 765–777. https://doi.org/10.1101/SQB.1990.055.01.072

Leibowitz, H., Brislin, R., Perlmutrer, L., & Hennessy, R. (1969). Ponzo Perspective Illusion as a Manifestation of Space Perception. *Science*, *166*(3909), 1174–1176. https://doi.org/10.1126/science.166.3909.1174

Leibowitz, H., & Moore, D. (1966). Role of changes in accommodation and convergence in the perception of size. *Journal of the Optical Society of America*, *56*(8), 1120–1129. https://doi.org/10.1364/JOSA.56.001120

Leibowitz, H. W., Shiina, K., & Hennessy, R. T. (1972). Oculomotor adjustments and size constancy. *Perception & Psychophysics*, *12*(6), 497–500. https://doi.org/10.3758/BF03210943

Li, Z., & Durgin, F. H. (2012a). A Comparison of Two Theories of Perceived Distance on the Ground Plane: The Angular Expansion Hypothesis and the Intrinsic Bias Hypothesis. *I-Perception*, *3*(5), 368–383. https://doi.org/10.1068/i0505

Li, Z., & Durgin, F. H. (2012b). A Comparison of Two Theories of Perceived Distance on the Ground Plane: The Angular Expansion Hypothesis and the Intrinsic Bias Hypothesis. *I-Perception*, *3*(5), 368–383. https://doi.org/10.1068/i0505

Lie, I. (1965). Convergence as a cue to perceived size and distance. *Scandinavian Journal of Psychology*, *6*(4), 109–116. https://doi.org/10.1111/j.1467-9450.1965.tb01059.x

Lindner, A., Iyer, A., Kagan, I., & Andersen, R. A. (2010). Human Posterior Parietal Cortex Plans Where to Reach and What to Avoid. *Journal of Neuroscience*, *30*(35), 11715–11725. https://doi.org/10.1523/JNEUROSCI.2849-09.2010

Linton, P. (in press). Does Vision Extract Absolute Distance from Vergence? *Attention, Perception, & Psychophysics*. Preprint: https://www.biorxiv.org/content/10.1101/731109v2

Linton, P. (2017). *The Perception and Cognition of Visual Space*. Palgrave.

Linton, P. (2018). *Do We See Scale?* https://www.biorxiv.org/content/10.1101/371948v1

Linton, Paul. (2019). *Would Gaze-Contingent Rendering Improve Depth Perception in Virtual and Augmented Reality?* https://arxiv.org/abs/1905.10366v1

Linton, Paul. (2020). Does vision extract absolute distance from vergence? *Attention, Perception, & Psychophysics*, *82*(6), 3176–3195. https://doi.org/10.3758/s13414-020-02006-1

Lisi, M. (2015). *Linear and generalized linear mixed-effects model in R (tutorial)*. https://mattelisi.github.io/files/mm_slides.pdf

Liu, S., Hua, H., & Cheng, D. (2010). A Novel Prototype for an Optical See-Through Head-Mounted Display with Addressable Focus Cues. *IEEE Transactions on Visualization and Computer Graphics*, *16*(3), 381–393. https://doi.org/10.1109/TVCG.2009.95

Locke, J. (1849). XXV. On single and double vision produced by viewing objects with both eyes; and on an optical illusion with regard to the distance of objects. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *34*(228), 195–201. https://doi.org/10.1080/14786444908646207

Locke, N. M. (1938). Some Factors in Size-Constancy. *The American Journal of Psychology*, *51*(3), 514–520. JSTOR. https://doi.org/10.2307/1416151

Loftus, A., Servos, P., Goodale, M. A., Mendarozqueta, N., & Mon-Williams, M. (2004). When two eyes are better than one in prehension: Monocular viewing and end-point variance. *Experimental Brain Research*, *158*(3), 317–327. https://doi.org/10.1007/s00221-004-1905-2

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, *293*, 133–135. https://doi.org/10.1038/293133a0

Loomis, J. M. (2001). Looking down is looking up. *Nature*, *414*(6860), 155–156. https://doi.org/10.1038/35102648

Lou, L. (2007). Apparent Afterimage Size, Emmert's Law, and Oculomotor Adjustment. *Perception*, *36*(8), 1214–1228. https://doi.org/10.1068/p5513

Lugtigheid, A. J., Brenner, E., & Welchman, A. E. (2011). Speed judgments of three-dimensional motion incorporate extraretinal information. *Journal of Vision*, *11*(13), 1–1. https://doi.org/10.1167/11.13.1

Lugtigheid, A. J., Wilcox, L. M., Allison, R. S., & Howard, I. P. (2014). Vergence eye movements are not essential for stereoscopic depth. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1776). https://doi.org/10.1098/rspb.2013.2118

Luneburg, R. K. (1947). *Mathematical analysis of binocular vision*. Princeton University Press.

Lunn, P. D., & Morgan, M. J. (1997). Discrimination of the spatial derivatives of horizontal binocular disparity. *JOSA A*, *14*(2), 360–371. https://doi.org/10.1364/JOSAA.14.000360

Luo, X., Kenyon, R., Kamper, D., Sandin, D., & DeFanti, T. (2007). The Effects of Scene Complexity, Stereovision, and Motion Parallax on Size Constancy in a Virtual Environment. *2007 IEEE Virtual Reality Conference*, 59–66. https://doi.org/10.1109/VR.2007.352464

Luo, Xun, V. Kenyon, R., G. Kamper, D., J. Sandin, D., & A. DeFanti, T. (2015). On the Determinants of Size-Constancy in a Virtual Environment. *International Journal of Virtual Reality*, *08*(1), 43–51.

Marin, F., Rohatgi, A., & Charlot, S. (2017). WebPlotDigitizer, a polyvalent and free software to extract spectra from old astronomical publications: Application to ultraviolet spectropolarimetry. *ArXiv:1708.02025 [Astro-Ph]*. http://arxiv.org/abs/1708.02025

Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature*, *389*(6648), 283–286. https://doi.org/10.1038/38496

Mayhew, J. E. W., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, *297*(5865), 376–378. https://doi.org/10.1038/297376a0

McCready, D. W. (1965). Size-distance perception and accommodation-convergence micropsia—A critique. *Vision Research*, *5*(4), 189–206. https://doi.org/10.1016/0042-6989(65)90065-9

Mckee, S. P., & Welch, L. (1992). The precision of size constancy. *Vision Research*, *32*(8), 1447–1460. https://doi.org/10.1016/0042-6989(92)90201-S

Melmoth, D. R., & Grant, S. (2006). Advantages of binocular vision for the control of reaching and grasping. *Experimental Brain Research*, *171*(3), 371–388. https://doi.org/10.1007/s00221-005-0273-x

Melmoth, D. R., Storoni, M., Todd, G., Finlay, A. L., & Grant, S. (2007). Dissociation between vergence and binocular disparity cues in the control of prehension. *Experimental Brain Research*, *183*(3), 283–298. https://doi.org/10.1007/s00221-007-1041-x

Metlapally, S., Tong, J. L., Tahir, H. J., & Schor, C. M. (2014). The impact of higher-order aberrations on the strength of directional signals produced by accommodative microfluctuations. *Journal of Vision*, *14*(12). https://doi.org/10.1167/14.12.25

Metlapally, S., Tong, J. L., Tahir, H. J., & Schor, C. M. (2016). Potential role for microfluctuations as a temporal directional cue to accommodation. *Journal of Vision*, *16*(6). https://doi.org/10.1167/16.6.19

Meyer, H. (1842). Ueber einige Täuschungen in der Entfernung u. Grösse der Gesichtsobjecte. *Archiv Für Physiologische Heilkunde*.

Meyer, H. (1852). Ueber die Schätzung der Grösse und der Entfernung der Gesichtsobjecte aus der Convergenz der Augenaxen. *Annalen Der Physik*, *161*(2), 198–207. https://doi.org/10.1002/andp.18521610204

Millard, A. S., Sperandio, I., & Chouinard, P. A. (2020). The contribution of stereopsis in Emmert's law. *Experimental Brain Research*. https://doi.org/10.1007/s00221-020-05747-5

Mon-Williams, M., & Dijkerman, H. C. (1999). The use of vergence information in the programming of prehension. *Experimental Brain Research*, *128*(4), 578–582. https://doi.org/10.1007/s002210050885

Mon-Williams, M., & Tresilian, J. R. (1999). Some Recent Studies on the Extraretinal Contribution to Distance Perception. *Perception*, *28*(2), 167–181. https://doi.org/10.1068/p2737

Mon-Williams, M., & Tresilian, J. R. (2000). Ordinal depth information from accommodation? *Ergonomics*, *43*(3), 391–404. https://doi.org/10.1080/001401300184486

Mon-Williams, M., Tresilian, J. R., McIntosh, R. D., & Milner, A. D. (2001). Monocular and binocular distance cues: Insights from visual form agnosia I (of III). *Experimental Brain Research*, *139*(2), 127–136.

Mon-Williams, M., Tresilian, J. R., Plooy, A., Wann, J. P., & Broerse, J. (1997). Looking at the task in hand: Vergence eye movements and perceived size. *Experimental Brain Research*, *117*(3), 501–506. https://doi.org/10.1007/s002210050246

Morgan, M. J., Watamaniuk, S. N., & McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research*, *40*(17), 2341–2349. https://doi.org/10.1016/s0042-6989(00)00093-6

Morgan, Michael J. (1989). Vision of solid objects. *Nature*, *339*(6220), 101–103. https://doi.org/10.1038/339101a0

Morrison, J. D., & Whiteside, T. C. (1984). Binocular cues in the perception of distance of a point source of light. *Perception*, *13*(5), 555–566. https://doi.org/10.1068/p130555

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*(3), 429–434. https://doi.org/10.1038/nn1641

Naceri, A., Chellali, R., & Hoinville, T. (2011). Depth Perception Within Peripersonal Space Using Head-Mounted Display. *Presence: Teleoperators and Virtual Environments*, *20*(3), 254–272. https://doi.org/10.1162/PRES_a_00048

Naceri, A., Moscatelli, A., & Chellali, R. (2015). Depth discrimination of constant angular size stimuli in action space: Role of accommodation and convergence cues. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00511

Nawrot, M., Ratzlaff, M., Leonard, Z., & Stroyan, K. (2014). Modeling depth from motion parallax with the motion/pursuit ratio. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01103

Nefs, H. T., & Harris, J. M. (2008). Induced motion in depth and the effects of vergence eye movements. *Journal of Vision*, *8*(3), 8–8. https://doi.org/10.1167/8.3.8

Neggers, S. F., & Bekkering, H. (1999). Integration of visual and somatosensory target information in goal-directed eye and arm movements. *Experimental Brain Research*, *125*(1), 97–107. https://doi.org/10.1007/s002210050663

Newman, D. G., & Ostler, D. (2009, July). *Hyperstereopsis Associated with Helmet-Mounted Sighting and Display Systems for Helicopter Pilots* [Text]. Aerospace Medical Association. https://doi.org/info:doi/10.3357/ASEM.21005.2009

Ogle, K. N. (1953). Precision and Validity of Stereoscopic Depth Perception from Double Images*. *JOSA*, *43*(10), 906–913. https://doi.org/10.1364/JOSA.43.000906

Olson, T. J., & Coombs, D. J. (1991). Real-time vergence control for binocular robots. *International Journal of Computer Vision*, *7*(1), 67–89. https://doi.org/10.1007/BF00130490

Ono, H., & Comerford, J. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and process*. Wiley.

Ono, H., Mitson, L., & Seabrook, K. (1971). Change in convergence and retinal disparities as an explanation for the wallpaper phenomenon. *Journal of Experimental Psychology*, *91*, 1–10. https://doi.org/10.1037/h0031795

Ooi, T. L., & He, Z. J. (2015). Space Perception of Strabismic Observers in the Real World Environment. *Investigative Ophthalmology & Visual Science*, *56*(3), 1761–1768. https://doi.org/10.1167/iovs.14-15741

Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature*, *414*(6860), 197–200. https://doi.org/10.1038/35102562

Owens, D. A., & Liebowitz, H. W. (1980). Accommodation, convergence, and distance perception in low illumination. *American Journal of Optometry and Physiological Optics*, *57*(9), 540–550. https://doi.org/10.1097/00006324-198009000-00004

Parker, A. J., Smith, J. E. T., & Krug, K. (2016). Neural architectures for stereo vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1697), 20150261. https://doi.org/10.1098/rstb.2015.0261

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*. https://doi.org/10.3389/neuro.11.010.2008

Pinheiro, J., & Bates, D. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag. //www.springer.com/gb/book/9780387989570

Pisella, L., Gréa, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., Boisson, D., & Rossetti, Y. (2000). An “automatic pilot” for the hand in human posterior parietal cortex: Toward reinterpreting optic ataxia. *Nature Neuroscience*, *3*(7), 729–736. https://doi.org/10.1038/76694

Poulton, E. C. (1980). Human manual control. In V. B. Brooks (Ed.), *Hanbook of Physiology 2: The nervous system* (pp. 1337–1389). American Physiological Association.

Poulton, E. C. (1988). The Journal of Motor Behavior in the 1960s and the 1980s. *Journal of Motor Behavior*, *20*(1), 75–78. https://doi.org/10.1080/00222895.1988.10735434

Predebon, J. (1979). Effect of familiar size on judgments of relative size and distance. *Perceptual and Motor Skills*, *48*(3 Pt 2), 1211–1214. https://doi.org/10.2466/pms.1979.48.3c.1211

Predebon, John. (1987). Familiar size and judgments of distance: Effects of response mode. *Bulletin of the Psychonomic Society*, *25*(4), 244–246. https://doi.org/10.3758/BF03330344

Predebon, John. (1990). Relative distance judgments of familiar and unfamiliar objects viewed under representatively natural conditions. *Perception & Psychophysics*, *47*(4), 342–348. https://doi.org/10.3758/BF03210874

Predebon, John. (1992a). The Influence of Object Familiarity on Magnitude Estimates of Apparent Size. *Perception*, *21*(1), 77–90. https://doi.org/10.1068/p210077

Predebon, John. (1992b). The role of instructions and familiar size in absolute judgments of size and distance. *Perception & Psychophysics*, *51*(4), 344–354. https://doi.org/10.3758/BF03211628

Predebon, John. (1993). The Familiar-Size Cue to Distance and Stereoscopic Depth Perception. *Perception*, *22*(8), 985–995. https://doi.org/10.1068/p220985

Predebon, John. (1994). Perceived size of familiar objects and the theory of off-sized perceptions. *Perception & Psychophysics*, *56*(2), 238–247. https://doi.org/10.3758/BF03213902

Predebon, John, & Woolley, J. S. (1994). The Familiar-Size Cue to Depth under Reduced-Cue Viewing Conditions. *Perception*, *23*(11), 1301–1312. https://doi.org/10.1068/p231301

Priestley, J. (1772). *The History and Present State of Discoveries Relating to Vision, Light, and Colours*. J. Johnson.

Prins, N., & Kingdom, F. A. A. (2018). Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01250

Prins, N., & Kingdom, F. A. A. (2019). http://www.palamedestoolbox.org/hierarchicalbayesian.html

Priot, A. E., Neveu, P., Philippe, M., & Roumes, C. (2012). Adaptation to alterations of apparent distance in stereoscopic displays: From lab to hyperstereopsis. *2012 International Conference on 3D Imaging (IC3D)*, 1–7. https://doi.org/10.1109/IC3D.2012.6615123

Priot, A.-E., Laboissière, R., Plantier, J., Prablanc, C., & Roumes, C. (2011). Partitioning the components of visuomotor adaptation to prism-altered distance. *Neuropsychologia*, *49*(3), 498–506. https://doi.org/10.1016/j.neuropsychologia.2010.11.028

Priot, A.-E., Laboissière, R., Sillan, O., Roumes, C., & Prablanc, C. (2010). Adaptation of egocentric distance perception under telestereoscopic viewing within reaching space. *Experimental Brain Research*, *202*(4), 825–836. https://doi.org/10.1007/s00221-010-2188-4

Priot, A.-E., Vacher, A., Vienne, C., Neveu, P., & Roumes, C. (2018). The initial effects of hyperstereopsis on visual perception in helicopter pilots flying with see-through helmet-mounted displays. *Displays*, *51*, 1–8. https://doi.org/10.1016/j.displa.2017.11.002

Quinlan, D. J., & Culham, J. C. (2007). FMRI reveals a preference for near viewing in the human parieto-occipital cortex. *NeuroImage*, *36*(1), 167–187. https://doi.org/10.1016/j.neuroimage.2007.02.029

R Core Team. (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. http://www.R-project.org/

Ramsay, A. I. G., Carey, D. P., & Jackson, S. R. (2007). Visual-proprioceptive mismatch and the Taylor illusion. *Experimental Brain Research*, *176*(1), 173–181. https://doi.org/10.1007/s00221-006-0609-1

Rand, K. M., Tarampi, M. R., Creem-Regehr, S. H., & Thompson, W. B. (2011). The Importance of a Visual Horizon for Distance Judgments under Severely Degraded Vision. *Perception*, *40*(2), 143–154.

Rauschecker, A. M., Solomon, S. G., & Glennerster, A. (2006). Stereo and motion parallax cues in human 3D vision: Can they vanish without a trace? *Journal of Vision*, *6*(12), 12–12. https://doi.org/10.1167/6.12.12

Regan, D., Erkelens, C. J., & Collewijn, H. (1986). Necessary conditions for the perception of motion in depth. *Investigative Ophthalmology & Visual Science*, *27*(4), 584–597.

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013a). The perception of egocentric distances in virtual environments—A review. *ACM Computing Surveys*, *46*(2), 1–40. https://doi.org/10.1145/2543581.2543590

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013b). The Perception of Egocentric Distances in Virtual Environments—A Review. *ACM Comput. Surv.*, *46*(2), 23:1–23:40. https://doi.org/10.1145/2543581.2543590

Richards, W. (1968). Spatial remapping in the primate visual system. *Kybernetik*, *4*(4), 146–156. https://doi.org/10.1007/BF00288548

Rogers, B. (2009). Are stereoscopic cues ignored in telestereoscopic viewing? *Journal of Vision*, *9*(8), 288–288. https://doi.org/10.1167/9.8.288

Rogers, B. (2017). *Perception: A Very Short Introduction*. Oxford University Press.

Rogers, B. (2019). Toward a new theory of stereopsis: A critique of Vishwanath (2014). *Psychological Review*, *126*(1), 162–169. https://doi.org/10.1037/rev0000131

Rogers, B. J. (2006). Failures of stereoscopic depth constancy: Fact or artefact? [Abstract]. *Journal of Vision*, *6*(6), 271, 271a. https://doi.org/10.1167/6.6.271

Rogers, B. J. (2011). Information, illusion, and constancy in telestereoscopic viewing. In L. R. Harris & M. R. M. Jenkin, *Vision in 3D Environments*. Cambridge University Press.

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity Scaling and the Perception of Frontoparallel Surfaces. *Perception*, *24*(2), 155–179. https://doi.org/10.1068/p240155

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Sacks, O. (2006). Stereo Sue. *The New Yorker*, 64.

Scarfe, P., & Hibbard, P. (2017). A Bayesian model of distance perception from ocular convergence. *Journal of Vision*, *17*(10), 159. https://doi.org/10.1167/17.10.159

Scarfe, P., & Hibbard, P. B. (2013). Reverse correlation reveals how observers sample visual information when estimating three-dimensional shape. *Vision Research*, *86*, 115–127. https://doi.org/10.1016/j.visres.2013.04.016

Schechner, Y. Y., & Kiryati, N. (2000). Depth from Defocus vs. Stereo: How Different Really Are They? *International Journal of Computer Vision*, *39*(2), 141–162. https://doi.org/10.1023/A:1008175127327

Schnall, S. (2017). Social and Contextual Constraints on Embodied Perception. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(2), 325–340. https://doi.org/10.1177/1745691616660199

Schor, C. M., & Tyler, C. W. (1981). Spatio-temporal properties of Panum's fusional area. *Vision Research*, *21*(5), 683–692. https://doi.org/10.1016/0042-6989(81)90076-6

Schor, C., Wood, I., & Ogawa, J. (1984). Binocular sensory fusion is limited by spatial resolution. *Vision Research*, *24*(7), 661–665. https://doi.org/10.1016/0042-6989(84)90207-4

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2017). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*(1), 205–233.

Sedgwick, H. A. (1983). Environment-centered representation of spatial layout: Available visual information from texture and perspective. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision.* (pp. 425–458). Academic Press.

Shaffer, D. M., McManama, E., Swank, C., & Durgin, F. H. (2013). Sugar and space? Not the case: Effects of low blood glucose on slant estimation are mediated by beliefs. *I-Perception*, *4*(3), 147–155. https://doi.org/10.1068/i0592

Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change Blindness in the Absence of a Visual Disruption. *Perception*, *29*(10), 1143–1154. https://doi.org/10.1068/p3104

Smith, R. (1738). *A Compleat System of Opticks in Four Books, Viz. A Popular, a Mathematical, a Mechanical, and a Philosophical Treatise. To which are Added Remarks Upon the Whole. By Robert Smith*. author, and sold there.

Sobel Erik C., & Collett Thomas Stephen. (1991). Does vertical disparity scale the perception of stereoscopic depth? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *244*(1310), 87–90. https://doi.org/10.1098/rspb.1991.0055

Sperandio, I., & Chouinard, P. A. (2015). The Mechanisms of Size Constancy. *Multisensory Research*, *28*(3–4), 253–283. https://doi.org/10.1163/22134808-00002483

Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nature Neuroscience*, *15*(4), 540–542. https://doi.org/10.1038/nn.3069

Sperandio, I., Kaderali, S., Chouinard, P. A., Frey, J., & Goodale, M. A. (2013). Perceived size change induced by nonvisual signals in darkness: The relative contribution of vergence and proprioception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(43), 16915–16923. https://doi.org/10.1523/JNEUROSCI.0977-13.2013

Stuart, G., Jennings, S., Kalich, M., Rash, C., Harding, T., & Craig, G. (2009). Flight performance using a hyperstereo helmet-mounted display: Adaptation to hyperstereopsis. *Proceedings of SPIE - The International Society for Optical Engineering*, *7326*. https://doi.org/10.1117/12.820500

Suzuki, K. (1986). Effects of oculomotor cues on the apparent size of afterimages. *Japanese Psychological Research*, *28*(4), 168–175. https://doi.org/10.4992/psycholres1954.28.168

Svarverud, E., Gilson, S., & Glennerster, A. (2012). A Demonstration of ‘Broken’ Visual Space. *PLOS ONE*, *7*(3), e33782. https://doi.org/10.1371/journal.pone.0033782

Svarverud, E., Gilson, S. J., & Glennerster, A. (2010). Cue combination for 3D location judgements. *Journal of Vision*, *10*(1), 5.1-513. https://doi.org/10.1167/10.1.5

Swan, J. E., Singh, G., & Ellis, S. R. (2015). Matching and Reaching Depth Judgments with Real and Augmented Reality Targets. *IEEE Transactions on Visualization and Computer Graphics*, *21*(11), 1289–1298. https://doi.org/10.1109/TVCG.2015.2459895

Swanston, M. T., & Gogel, W. C. (1986). Perceived size and motion in depth from optical expansion. *Perception & Psychophysics*, *39*(5), 309–326. https://doi.org/10.3758/BF03202998

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. https://doi.org/10.1080/00221309.1932.9918473

Taylor, F. V. (1941). Change in size of the afterimage induced in total darkness. *Journal of Experimental Psychology*, *29*(1), 75–80. https://doi.org/10.1037/h0058125

Thompson, W., Fleming, R., Creem-Regehr, S., & Stefanucci, J. K. (2011). *Visual Perception from a Computer Graphics Perspective*. CRC Press.

Török, Á., Ferrè, E. R., Kokkinara, E., Csépe, V., Swapp, D., & Haggard, P. (2017). Up, Down, Near, Far: An Online Vestibular Contribution to Distance Judgement. *PLOS ONE*, *12*(1), e0169990. https://doi.org/10.1371/journal.pone.0169990

Tresilian, J. R., Mon-Williams, M., & Kelly, B. M. (1999). Increasing confidence in vergence as a cue to distance. *Proceedings of the Royal Society of London B: Biological Sciences*, *266*(1414), 39–44. https://doi.org/10.1098/rspb.1999.0601

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., & Imbert, M. (1992). Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science*, *257*(5074), 1279–1281. https://doi.org/10.1126/science.1519066

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., & Imbert, M. (1996). Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. *Journal of Neurophysiology*, *76*(5), 2872–2885. https://doi.org/10.1152/jn.1996.76.5.2872

Trotter, Yves, & Celebrini, S. (1999). Trotter, Y. & Celebrini, S. Gaze direction controls response gain in primary visual-cortex neurons. Nature 398, 239−242. *Nature*, *398*, 239–242. https://doi.org/10.1038/18444

Trotter, Yves, Stricanne, B., Celebrini, S., Thorpe, S., & Imbert, M. (1993). *Neural processing of stereopsis as a function of viewing distance*. Oxford University Press. https://0-www-oxfordscholarship-com.wam.city.ac.uk/view/10.1093/acprof:oso/9780198547853.001.0001/acprof-9780198547853-chapter-027

Tyler, C. (2018). Absolute Distance From Perspective. *Perception*, *47*(6), 581–584. https://doi.org/10.1177/0301006618774625

Tyler, C. W. (1971). Stereoscopic Depth Movement: Two Eyes Less Sensitive than One. *Science*, *174*(4012), 958–961. https://doi.org/10.1126/science.174.4012.958

Urist, M. J. (1959). Afterimages and ocular muscle proprioception. *A.M.A. Archives of Ophthalmology*, *61*(2), 230–232. https://doi.org/10.1001/archopht.1959.00940090232006

van Donkelaar, P., & Staub, J. (2000). Eye-hand coordination to visual versus remembered targets. *Experimental Brain Research*, *133*(3), 414–418. https://doi.org/10.1007/s002210000422

Vesia, M., Prime, S. L., Yan, X., Sergio, L. E., & Crawford, J. D. (2010). Specificity of Human Parietal Saccade and Reach Regions during Transcranial Magnetic Stimulation. *Journal of Neuroscience*, *30*(39), 13053–13065. https://doi.org/10.1523/JNEUROSCI.1644-10.2010

Viguier, A., Clément, G., & Trotter, Y. (2001). Distance perception within near visual space. *Perception*, *30*(1), 115–124. https://doi.org/10.1068/p3119

Vishwanath, D. (2019). Advancing a New Theory of Stereopsis: Reply to Rogers (2019). *Psychological Review*.

Vishwanath, Dhanraj. (2010). Reconciling Pictures and Reality: Information in Surface and Depth Perception. In L. Albertazzi, G. J. van Tonder, & D. Vishwanath (Eds.), *Perception Beyond Inference: The Information Content of Visual Processes* (pp. 201–240). MIT Press.

Vishwanath, Dhanraj. (2014). Toward a new theory of stereopsis. *Psychological Review*, *121*(2), 151–178. https://doi.org/10.1037/a0035233

Vishwanath, Dhanraj, & Blaser, E. (2010). Retinal blur and the perception of egocentric distance. *Journal of Vision*, *10*(10), 26–26. https://doi.org/10.1167/10.10.26

Von Hofsten, C. (1976). The role of convergence in visual space perception. *Vision Research*, *16*(2), 193–198. https://doi.org/10.1016/0042-6989(76)90098-5

Von Holst, E. (1955b). 1st der Einfluss der Akkommodation auf die gesehene Dinggrijsse ein “reflektorischer” Vorgang? *Naturwissenschaften*, *42*, 445–446.

Von Holst, E. (1955a). Die Beteiligung von Konvergenz und Akkommodation an der wahrgenommenen Grossenkonstanz. *Naturwissenschaften*, *42*, 444–445.

Von Holst, E. (1957). Aktive Leistungen der menschlichen Gesichtswahrnemung. *Studium Generale*, *10*, 231–243.

Wallach, H., Gillam, B., & Cardillo, L. (1979). Some consequences of stereoscopic depth constancy. *Perception & Psychophysics*, *26*(3), 235–240. https://doi.org/10.3758/BF03199874

Wallach, H., & Zuckerman, C. (1963). The constancy of stereoscopic depth. *The American Journal of Psychology*, *76*, 404–412.

Walls, G. L. (Gordon L. (1942). *The vertebrate eye and its adaptive radiation*. Bloomfield Hills, Mich., Cranbrook Institute of Science. http://archive.org/details/vertebrateeyeits00wall

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, *17*(3), 10–10. https://doi.org/10.1167/17.3.10

Welchman, A. E., Harris, J. M., & Brenner, E. (2009). Extra-retinal signals support the estimation of 3D motion. *Vision Research*, *49*(7), 782–789. https://doi.org/10.1016/j.visres.2009.02.014

Wheatstone, C. (1852). I. The Bakerian Lecture.—Contributions to the physiology of vision.—Part the second. On some remarkable, and hitherto unobserved, phenomena of binocular vision (continued). *Philosophical Transactions of the Royal Society of London*, *142*, 1–17. https://doi.org/10.1098/rstl.1852.0001

Wildenmann, U., & Schaeffel, F. (2013). Variations of pupil centration and their effects on video eye tracking. *Ophthalmic and Physiological Optics*, *33*(6), 634–641. https://doi.org/10.1111/opo.12086

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., & Merfeld, D. M. (2019). *Sensation and Perception (International Edition)*. Sinauer Associates, Oxford University Press.

Wundt, W. M. (1862). *Beiträge zur theorie der sinneswahrnehmung*. C. F. Winter. http://archive.org/details/beitrgezurtheor00wundgoog

Wyatt, H. J. (2010). The human pupil and the use of video-based eyetrackers. *Vision Research*, *50*(19), 1982–1988. https://doi.org/10.1016/j.visres.2010.07.008

Zenkin, G. M., & Petrov, A. P. (2015). Transformation of the Visual Afterimage Under Subject's Eye and Body Movements and the Visual Field Constancy Mechanisms. *Perception*, *44*(8–9), 973–985. https://doi.org/10.1177/0301006615594937

Ziegler, L. R., & Hess, R. F. (1997). Depth perception during diplopia is direct. *Perception*, *26*(10), 1125–1130.