Title:

Advancing a new theory of stereopsis: Reply to Rogers (2019)

Author:

Dhanraj Vishwanath

Affiliation:

School of Psychology & Neuroscience

University of St. Andrews, St. Andrews, Scotland, UK

Email: dv10@st-andrews.ac.uk

Abstract

Vishwanath (2014) presented analyses and proposed conjectures aimed at a unified understanding of both qualitative and quantitative aspects of stereopsis in pictorial and natural (real-world) 3D vision. A recent commentary by Rogers (2019) concedes the key argument in the paper, that stereopsis can be induced in the absence of binocular disparity and motion parallax, but criticizes the wider analyses and conjectures. Rogers argues that a focus on visual appearance and qualitative aspects of 3D perception is unproductive and that the analysis of pictorial space perception adds little to our wider understanding of 3D vision. I argue here that the critique is not persuasive as it misconstrues the distinction between qualitative and quantitative aspects of perception and its claims regarding pictorial depth perception rely on introspections that often do not align with the empirical record. I reaffirm that an integrative focus on both qualitative and quantitative aspects of both pictorial and natural 3D perception is crucial for advancing an understanding of the complex phenomenon of stereopsis.

Keywords

Stereopsis; 3D; picture perception; depth perception; cue conflict; binocular disparity

Introduction

In "Eye, Brain, Vision", Nobel laureate David Hubel commented on how remarkable it was that the physiologist Ewald Hering deduced cell-level central-nervous-system antagonistic color mechanisms (opponent cells) through the analysis of purely subjective (qualitative) aspects of colour appearance, at a time when very little was understood in neurophysiology. Hubel highlights how, despite this achievement, analysis of qualitative aspects of perception is still viewed with scepticism in vision science, saying "to the extent that I am slightly to the Hering side of center, I will doubtless make enemies of all the experts". Nearly five years ago, inspired by Hering's approach, I presented analyses and theoretical conjectures aimed at a unified understanding of qualitative and quantitative aspects of human 3D vision (Vishwanath, 2014). Brian Rogers, an expert in 3D vision, has criticised these analyses and proposals (Rogers, 2019)

Rogers argues that analysing qualitative aspects of 3D perception is unproductive and that aiming to understand stereopsis in terms of "visual appearance" erroneously construes stereopsis to be a merely "qualitative" phenomenon. Rogers claims that stereopsis (and natural 3D vision in general) is the perception of "quantitative depth".

These arguments suggest a misunderstanding of the distinction between "visual appearance" on one hand, and qualitative and quantitative aspects of visual appearance on the other. Importantly, they miss the point that the terms "qualitative" and "quantitative" with respect to perception are strictly operationalizations and are not mutually exclusive constructs. It is useful therefore to highlight this distinction.

Quantitative and qualitative aspects of visual appearance

The analysis in Vishwanath (2014) is based on the premise that what is fundamentally given in visual perception are subjective visual presentations or appearances; and that it is on the basis of these “appearances” that we operationally define quantitative or qualitative attributes of perception.

For example, if asked to look around the room you are in, I’m sure you would agree that you have a visual presentation of objects and space before you. There is nothing explicitly “quantitative” about any aspect of this visual presentation. There are no Post-It notes in your visual field with numbers on them specifying sizes of objects, distances between them, or their position with respect you, the observer. We label an attribute of visual appearance to be “quantitative” when we understand it to have a simple correlation to some measurable physical property and when it seems straightforward to make quantitative magnitude judgements about it. For example, I might perceive one side of the book on my desk to be brighter than the other side (which is in shadow). I might perceive the book to be closer to the window than the window is to the tree beyond, or, perceive the book to be narrower than another book on my desk. For these attributes of visual appearance (brightness, distance and shape) there is a readily verifiable correlation with measurable physical properties: brightness relates to quantity of impinging light and shape or distance relate to lengths measured with a ruler. Also, there is a straightforward way to make a quantitative judgement about these attributes of appearance. For example, I might report, on a 1-10 scale, that the brightness of the left side of the book is a “5” while the part in the shadow is a “3”; that one book has an aspect ratio of 1.5 while the other, 1.0; that the tree is twice as far from the window as the window is from the book.

Other attributes of the visual appearance are not so easily quantifiable or relatable to physical magnitudes. For example, I might notice that the chromatic appearance of the blue book on my desk is different from the red book, and different from the green leaves of the tree beyond the window. There is no straightforward way to define this aspect of appearance in terms of quantitative magnitudes, nor a simple quantitative correlation between chromatic appearance and a relevant physical property (light intensity or wavelength). Similarly, in the spatial domain, I might notice that that the leaves and branches of the tree outside appear to be in vivid depth relief (impression of stereopsis) when viewed with two eyes and less so with one eye. Here, like chromatic appearance, there is no straightforward way to ascribe quantitative magnitudes to this change in appearance, nor identify a correlation with a change in a relevant physical property. Clearly, the tree has not changed its shape or physical dimensions.

We might, as a starting point, characterize these latter aspects of visual appearance [chromatic appearance and spatial vividness (a.k.a., stereopsis)] as “qualitative attributes” of perception.

By paying careful attention to how these so-called qualitative attributes vary across stimulus conditions we can develop ways to make them more concrete or “quantifiable”. For example, we might realise that chromatic appearance can be broken down into two further attributes, hue and saturation. We can develop ways to “quantify” saturation by defining an arbitrary continuous scale from pure grey to a (say) pure blue. Similarly, we can “quantify” the qualitative attribute of spatial vividness by developing an arbitrary quantitative scale (Vishwanath & Hibbard, 2103) and a model linking changes in this attribute to changes in aspects of physical stimulation, and their encoding (Vishwanath, 2014).

However, the fact that an attribute of visual appearance is amenable to quantitative judgement, or correlates simply with measurable physical properties, does not make it a “quantitative perception” as implied by Rogers introspections that he perceives “quantitative depth” when stereopsis is present (e.g., real scenes), but not when he looks into pictorial space.

> “I do not perceive the [object] to stand out (i.e., to have quantitative depth) from the surface of the photo” (p. 163)

Here, it is hard to understand Rogers’ linkage of an object’s appearance of “standing out” to a notion of “quantitative depth”. The only plausible interpretation of “quantitative depth” is as an abbreviation for “an aspect of visual appearance which affords quantitative judgements of depth”. Such quantitative judgements may be explicit (reporting the distance, shape or size of an object in a psychophysical experiment) or implicit (reaching out and opening one’s fingers by the correct amount to pick up a mug). Accepting this more sensible definition, Rogers’ introspective claim that he does not perceive “quantitative depth” in pictorial images, such as Figure 1, becomes problematic.

> “the one thing that I do not perceive when looking at a picture or photograph is a sense of ‘quantitative’ depth of the objects depicted in the scene” (p. 163)

This is because there is widespread empirical evidence showing that naïve observers can make stable quantitative judgements of 3-D properties including object size, distance, surface slant, 3D shape and 3D layout when viewing pictures (e.g., Cooper & Banks, 2012; Erkelens, 2015, Held et al., 2010; Vishwanath et al., 2005, Koenderink et al., 1994, Koenderink et al., 2001; Kubovy, 1986; Wijntjes & Pont, 2012), just as they can when viewing real scenes and objects.

In contrast to Rogers (2019), Vishwanath (2014) proposed that both stereopsis and pictorial depth are types of perception where there is a visual *appearance* of 3D objects and space, but that they differ in both qualitative and quantitative aspects. The distinction between the two in *qualitative* terms is that under stereopsis, but not pictorial depth perception, there is an impression of negative space, object solidity, tangibility (what Rogers refers to as perceiving the object to “stand out”). The distinction between the two in *quantitative* terms is that pictures afford only quantitative judgements of *relative* spatial attributes (layout, object shape, relative size, etc.), while in conditions where stereopsis is present, judgements of *absolute* (scaled) spatial attributes (egocentric distance, absolute depth and size) are also afforded.

Furthermore, Vishwanath (2014) proposed that the perceived strength of the “qualitative” impression of stereopsis (solidity, negative space and tangibility) is correlated with the precision with which judgements of scaled spatial attributes (absolute size and depth) can be made. The qualitative aspect of stereopsis in the domain of 3D perception is analogous to the qualitative attribute of “saturation” in the domain of colour perception. Both attributes can, with the right operationalization, be quantified, and their variation with stimulus and viewing conditions related to underlying mechanisms and representations.

Rogers’ notion of stereopsis, which associates the impression of “standing out” simply with “quantitative depth”, provides no basis for empirical validation. In contrast, the proposal in Vishwanath (2014) is amenable to empirical validation through qualitative and quantitative psychophysics (Vishwanath & Hibbard, 2013; Hibbard, Hornsey & Haines, 2017), visuomotor interaction measures (Volcic et al, 2014) and neurophysiology (Uji et al., 2019a, 2019b)

On pictorial depth perception

Rogers (2019) devotes a large part of his argumentation to pictorial depth perception but simultaneously claims that an understanding of pictorial depth adds little to our understanding of natural (real-world) 3D vision.

> "How the visual system copes with unnatural and cue-conflict situations [pictures] may tell us rather little about how we perceive the structure and layout of the real 3D world". (p.163)

This is a puzzling statement. Not only is the analysis of pictorial depth perception responsible for the birth and development of the scientific question of human 3D vision and stereopsis (Wade, Ono & Lillakis, 2001; Wheatstone, 1938; Kubovy, 1986), the vast majority of 3D perception researchers acknowledged that an understanding of pictorial depth is part and parcel of a full understanding of human 3D vision (e.g., Cutting, 2003, Erkelens, 2015; Gibson, 1950; Koenderink, 1998; Kubovy, 1986; Pirenne, 1970; Sedgwick, 2003). Cue-conflict stimuli have also been used to delineate fundamental mechanisms underlying depth perception (Hillis et al., 2001; Ban et al. 2012).

Moreover, Rogers' discussion of picture perception seems contradictory. Rogers first quotes Gregory (1966) to highlight the commonly accepted observation that when viewing a picture of a 3D scene (e.g. Fig 1), one simultaneously perceives the "real" flat picture surface and also a "virtual" 3d space of the depicted pictorial scene. But Rogers' own introspections regarding pictures are hard to follow:

> With respect to the qualitative aspects of picture perception, there is nothing intangible or vague about my percept, nor is it lacking in solidity, what I perceive is the solid, tangible, flat surface of the picture, just as Gregory described it. (p. 164).

Rogers (2019) appears to be claiming that all that is perceived when looking at a picture of a 3D scene is the flat picture surface and nothing more. He claims he does not perceive any 3-dimensionality within the picture itself. Referring to the surfaces depicted in the picture, he says:

> "the surfaces themselves do not appear to be slanted, curved…" (p. 164)

But this goes against the self-evident fact that viewing a perspective image such as in Fig 1, generates perceptions of 3D objects and space (including slanted and curved surfaces), underlining why perspective images have been used to convey accurate perceptions of shape and layout by artists, architects and designers since the Renaissance. Rogers's statements align with a minority philosophical view that depth in pictures is simply a malleable cognitive interpretation or semiotic convention and not an actual perception; a view largely rejected by researchers in depth perception (see Hecht et al., 2003; Niederée & Hayer, 2003; Sedgwick, 2003).

In contrast, Vishwanath (2014) proposed that pictorial depth perception constitutes the perception of relative (unscaled) 3D structure while stereopsis constitutes the perception of absolute (egocentrically scaled) 3D structure, where scale is perceived with some degree of certainty. Correct-perspective pictorial images are valuable because they convey generally accurate perceptions of 3D shape and layout (relative depth structure) despite the ambiguity of scale. Moreover, the ambiguity of scale in pictures turns out to be beneficial. Regardless of the size of the pictorial image, we can malleably ascribe a visual scale to the pictorial objects consistent with recognizable familiar-sized objects while maintaining a stable and relatively accurate perception of object shape and layout (Fig 2).

On the dissociation of absolute and relative depth

Importantly, the model in Vishwanath (2014) relies on a proposed dissociation between representations of relative (unscaled) and absolute (egocentrically scaled) depth, a dissociation argued to underlie the duality of picture perception: the simultaneous awareness of the picture surface in real (scaled) space and the perception of a virtual (unscaled) pictorial space.

Rogers (2019) dismisses the possibility that relative and absolute encoding of depth structure are dissociable:

> It may be convenient to make a conceptual distinction between absolute distance and relative depth but it does not follow that they are kept separate in human perceptual processing. (p. 166)

This is a surprising statement that goes against widespread empirical results that reveal a clear dissociation between representations of relative (allocentric or exocentric) judgements and absolute (scaled or egocentric) judgements (e.g., Loomis et al. 1992; Campagnioli, Croom & Domini, 2017; Glennerster et al., 1996). Indeed, it contradicts Rogers' own previous arguments for distinct mechanisms of depth representations based on results comparing depth ratio (relative depth) judgements and absolute (scaled) depth judgements.

> "the best way to account for the difference is to assume that the visual system uses separate mechanisms to process disparity information in each case and we argue against the notion that a single "internal model" could be used to perform both tasks." (Glennerster, Rogers & Bradshaw, 1996, p. 3453).

Rogers' critique of Vishwanath's (2014) explanation of how picture duality and the emergence of monocular stereopsis can potentially be understood from the standpoint of the relative/absolute depth dissociation and reattribution of scaling information also suffers from Rogers' expressed difficulty in appreciating the important difference between pictorial depth cues and binocular disparity:

> So it is difficult to see how the depth cues of texture, shading, and perspective in pictures are different from the binocular disparities and parallax motions (p. 164).

Monocular cues such as perspective and shading are different from binocular disparity (and motion parallax) in that they specify relative 3D structure (shape and layout) independent of scaling. In contrast, disparities must mandatorily be scaled by distance information (vergence) in order to uniquely specify 3D shape and layout, since unscaled disparities do not provide unambiguous information about relative depth (they specify depth only up to an affine stretch; see Glennerster et al., 1996)[i].

In Vishwanath (2014), this distinction is proposed to underlie the fact that certain cues (disparity, parallax, vergence) mandatorily accrue to the egocentrically scaled depth map (causing the perception of a flat picture surface in "real" space) while the monocular cues which specify a conflicting 3D structure remain unscaled and accrue to a separate relative depth map (causing a perception of a "virtual" pictorial space).

Rogers' critique of how this explanation of picture duality can then explain monocular stereopsis as process of reattribution of residual distance scaling information to monocular cues also suffers from a misinterpretation of the informativeness of distance cues. For example, in countering the provisional explanation of stereopsis under synoptic viewing, where there is no binocular disparity and where I stated that the parallel vergence state is largely non-informative about distance, Rogers says the following:

> "This is incorrect. The vergence signal indicates viewing at a large distance…why should the eye vergence information, signaling that the objects in the scene are (and are seen to be at) a large distance away, be less effective in this assignment process than the much less precise accommodation signals?" (p. 167)

But a parallel vergence state is attained for any viewing distance between 6m and the farthest viewable object (e.g., the visual horizon or the moon); i.e., parallel vergence is largely non-informative (unreliable) with respect to the specification of distance and scale. In contrast, at the typical distances at which a picture is viewed with a synopter or monocular aperture (~0.5-2m), focus information driving accommodation is, in principle, quite informative (2-0.5 dioptres; Burge & Geisler, 2011).

On stereopsis in natural and stereoscopic vision.

Rogers (2019) claims that the analyses in Vishwanath (2014) are only applicable to picture perception and cue-conflict situations and do not help us understand depth perception in natural viewing. In fact, there were several important aspects of the variation in the appearance of stereopsis under natural viewing analysed in Vishwanath (2014). First, the subjective appearance of stereopsis (specifically, the impression of "real separation" in depth) was shown to be strongest for binocular viewing of objects in near (personal) space, diminishing with viewing distance independent of the magnitude of the depth separation between the viewed objects (Vishwanath, 2014, Fig 14). Related observations have been reported by others (Ogle, 1950; Tscherning, 1904; von Hildebrand, 1907). Second, as noted by da Vinci (quote below), monocular viewing of real scenes induces a degree of stereopsis (similar to that obtained under monocular viewing of pictorial images), but which also diminishes with distance such that objects at a distance viewed with one eye appear as though in a picture viewed with both eyes:

> "A Painting [viewed with both eyes]…can never show a Relievo equal to that of Natural Objects, unless these be view'd at a Distance and with a single Eye" [in Wade et al., 2001].

Vishwanath (2014) argued that a more unified explanation of these observations can be given by the fact that estimates of distance required for depth scaling are more reliable under binocular than monocular viewing in near space (vergence or successively scaled disparities) but become less reliable with increasing viewing distance (Cutting & Vishton, 1995, Loomis & Knapp, 2003, Palmisano et al., 2010). This can explain why binocular viewing induces a more vivid impression of stereopsis than monocular viewing in personal and action space, but a greater similarity in subjective appearance between the two (and to pictorial depth) is found in distant viewing (Vishwanath, 2014). Rogers (2019) notion of stereopsis as simply the perception of "quantitative depth" does not provide a basis for understanding any of these phenomena.

Regarding stereoscopic images, Rogers (2019) makes a claim that there is little difference in perceptual appearance of 3-dimensionality between viewing wide-field pictorial images (e.g., Imax™) and stereoscopic images:

> Perhaps this is one of the reasons why so-called 3D films (using disparate images) have not proved to be as popular as the filmmakers had hoped—the [binocular] disparities add very

> little to our perceptual experience compared with that obtained under large field viewing conditions [Imax] with little or no difference between the binocular images. (p164)

But this begs the question why directors of contemporary films [e.g., *Gravity* and *Avatar*] spent enormous sums of money and time devising complex stereoscopic rigs and processing pipelines to bring stereoscopic content (binocular disparity) to the viewers experience; despite having at their disposal the simpler solution of just filming in wide-field pictorial viewing (Imax). Contrary to Rogers' claim, Stereoscopic (3D) movies have waned in popularity, not because they "add very little to our perceptual experience", but because of a host of other well-documented factors, including the need for obtrusive eyewear, dimmer images due to image splitting, fatigue due to the accommodation-vergence conflict, frame violation, etc. (Banks et al., 2012). Rogers' claim confounds the trade-off inherent in stereoscopy (between generating a vivid perceptual experience and associated negative effects) with the question of under what conditions binocular disparity best adds to the visual experience of 3-dimensionality.

Advancing a new theory of stereopsis

Vishwanath (2014) provided an analysis of the viability of four existing definitions and hypotheses relating to stereopsis in comparison to the alternative proposed. The first was the "binocular disparity hypothesis", which is the mainstream definition of stereopsis as the perception of depth based on binocular disparities:

> "Depth perception based upon binocular disparities is known as stereopsis" (De Angelis, 2000)

The second (the “visual parallax hypothesis”), a variant of the mainstream definition, proposes that the impression of stereopsis is generated either from binocular disparities or monocularly in the presence of motion parallax (Rogers & Graham, 1989). Both these definitions must be disbanded if one accepts a critical observation: stereopsis can be induced under static monocular viewing of pictorial images or real scenes (i.e., monocular stereopsis). Rogers (2019) accepts the main claim in Vishwanath (2014) that stereopsis proper can be induced in the absence of both binocular disparity and motion parallax:

> “but would anyone want to claim that binocular disparities are necessary for stereopsis? I doubt it.” (p. 165)
>
> “there is no implication that visual parallax is the only source of information that could yield stereopsis. Thus, the “visual parallax hypothesis” is also a straw man” (p.165).

It is heartening that Rogers agrees with Vishwanath (2014) that these two views are no longer sustainable in light of empirical observation, though perhaps it is unfair to label such long-held hypotheses as “straw men”.

A third definition of stereopsis assessed in Vishwanath (2014) was one in the literature specifically aimed at explaining the phenomenon of monocular stereopsis. It was labelled the “cue-coherence/depth magnitude hypothesis” because it claims that stereopsis is induced when there is coherence among depth cues (e.g., viewing a real 3D scene) and not when there is conflict among them (viewing pictures binocularly). Specifically, it claims that when viewing pictures, binocular disparity (which specifies a flat surface) conflicts with and diminishes the depth specified by pictorial (monocular) cues. Monocular stereopsis (monocular aperture viewing, synoptic viewing, etc.) is proposed to arise due to the removal of the conflicting binocular information resulting in a more vivid qualitative impression of depth and a concomitant increase in the magnitude of perceived depth. Rogers dismisses this hypothesis because he believes it to be derived from mainstream cue-integration theory of depth perception

> “This seems to misrepresent both the findings and the modelling of cue-conflicts. As far as I am aware, no one working in this area would want to claim that there is correlation between the quantitative amount of slant or curvature and the qualitative impression of a “tangible solid form,”. (p. 165)

Rogers is correct that mainstream research in depth cue integration (e.g., Maloney & Landy, 1989) does not in any way address qualitative aspects of depth and 3-dimensionality. The “cue-coherence/depth magnitude hypothesis” instead represented proposals by other researchers who were interested in understanding the induction of the qualitative impression of stereopsis in single pictures (Ames, 1925; Koenderink, 1994; Schlosberg, 1948; Tscherning, 1900).

Despite taking issue with the cue-coherence/depth magnitude hypothesis, Rogers (2019) appeals to it as an explanation for the induction of monocular stereopsis (p. 164). However, he neglects to consider evidence that goes against a key prediction of this hypothesis. Empirical results from naïve observers show no difference in judged magnitude of depth comparing monocular and binocular viewing of pictures as predicted by this hypothesis (Cooper & Banks, 2012; Erkelens, 2015, Wijntjes & Pont, 2012; Vishwanath & Hibbard, 2013; c.f., Koenderink et al., 1994[ii]).

In summarizing his critique of my analysis of the four hypotheses for binocular and monocular stereopsis that predate Vishwanath (2014), Rogers (2019) states the following:

> The [4] differential predictions indicated in this table look much less impressive if (1) the first two hypotheses are ruled out as nonstarters, (2) [the] "cue-conflict/depth magnitude hypothesis" misrepresents the findings and predictions of cueconflict experiments, and (3) the "picture awareness hypothesis" represents just one aspect of cue-conflict. (p166)

Thus, Rogers and I concur in rejecting the four extant hypotheses and definitions of stereopsis that predate Vishwanath (2014). However, though their differential predictions, in hindsight, look "much less impressive", it would have been misleading not to assess them against those of the alternative proposed in Vishwanath (2014).

Conclusion

Rogers (2019) has criticized Vishwanath (2014) largely on the basis of two claims. The first is that visual appearance and subjective (qualitative) aspects of depth perception are not critical for the understanding human 3D vision. Yet almost all of Rogers' arguments are based on appealing to personal introspections regarding how depth and 3-dimensionality *appear*. The second is the claim that Vishwanath (2014) only pertains to pictorial depth perception rather than natural (real-world) perception, and that pictorial perception is unimportant for understanding human perception. Yet Rogers has dedicated his entire commentary to discuss only picture perception, providing no assessment or counter arguments on the detailed analyses provided in Vishwanath (2014) about stereopsis in natural (real-world) viewing. Moreover, the claims regarding picture perception, which rely almost exclusively on introspections, go against much of the empirical evidence and consensus understanding of the phenomena.

But there is a silver lining here. Rogers (2019) unequivocally accepts the central guiding assumption in Vishwanath (2014) that stereopsis proper can be induced in the absence of binocular disparity and visual parallax. This acceptance is the first step in recognising the need for advancing a new theory of the fundamental perceptual property that colours the appearance of 3-dimensional space: *stereopsis*.

References

Barry, S. (2009). *Fixing My Gaze.* New York: Basic Books

Ban H, Preston TJ, Meeson A & Welchman AE (2012) The integration of motion and disparity cues to depth in dorsal visual cortex. Nature Neuroscience, 15, 636-43,

Banks, M.S., Read, J.C.A., Allison, R.S., and Watt, S.J. (2012) Stereoscopy and the human visual system. *SMPTE Motion Imaging Journal*, 121 (4), 24-43.

Burge, J. & Geisler W.S. (2011). Optimal defocus estimation in individual natural images. *Proceedings of the National Academy of Sciences*, 108 (40): 16849-16854

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

Cutting, J. E., & Vishton, P. M. (1995). Perceiving Layout: The Integration, Relative Dominance, and Contextual Use of Different Information About Depth. In *Handbook of Perception and Cognition: Vol. 5: Perception of Space and Motion*, edited by W. Epstein and S. Rogers. New York: Academic Press. 69-117

Erkelens, C. J. (2015). The extent of visual space inferred from perspective angles. i-Perception, 6, 5–14.

DeAngelis, C. G. (2000) Seeing in three dimensions: the neurophysiology of stereopsis. *Trends in cognitive sciences* 4 (3), 80-90

Gogel, W.C. & Teitz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception & Psychophysics, 13*, 184-292.

Hecht, H., Schwartz, R., & Atherton, M. (2003) *Looking Into Pictures: An Interdisciplinary Approach to Pictorial Space.* Cambridge: The MIT Press.

Held, R., Cooper, E., O'Brien, J. & Banks, M. (2010). Using blur to affect distance and size. *ACM Transactions in Graphics*, 29(2):19, 1-16.

Hibbard, PB., Haines, AE. and Hornsey, RL., (2017). Magnitude, precision, and realism of depth perception in stereoscopic vision. *Cognitive Research: Principles and Implications.* 2 (25), 25

Hubel, D. H. (1988) *Eye*, *brain and vision*. W H Freeman, New York.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1994). On So-Called Paradoxical Monocular Stereoscopy. *Perception, 23*, 583–594.

Koenderink, J. J. (1998). Pictorial Relief. *Philosophical Transactions of the Royal Society London A, 356*, 1071–1086.

Kubovy, M. (1986). *The Psychology of Perspective and Renaissance Art.* New York: Cambridge University Press.

Loomis, J. M., Da Silva, J.A., Fujita, N., & Fukusima, S. S. (1992) Visual space perception and visually directed action. Journal of Experimental Psychology: Human Perception and Performance, 18, 906-921.

Loomis, J. M. & Knapp, J. M. (2003). Visual perception of egocentric distance in real and virtual environments. In L. J. Hettinger and M. W. Haas (Eds.), Virtual and Adaptive Environments (pp. 21-46). Mahwah NJ: Erlbaum

Maloney, L., & Landy, M. (1989). *A statistical framework for robust fusion of depth information*. *Proceedings of SPIE - The International Society for Optical Engineering* 1199

Michotte, A. (1991). The psychological enigma of perspective in outline pictures. In G. Thines, A. Costall, & G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception* (pp. 174–187). Hillsdale, NJ: Erlbaum. (Original work published 1948)

Niederée, R., & Heyer, D. (2003). The dual nature of picture perception: A challenge to current general accounts of visual perception. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), Looking into pictures: An interdisciplinary approach to pictorial space (pp. 77-98). Cambridge, MA, US: MIT Press.

Ogle, K. N. (1950). *Researches in Binocular Vision*. Philadelphia: W.B. Saunders

Palmisano, S.A., Gillam, B.J., Govan, D., Allison, R.S., & Harris, J.M. (2010). Stereoscopic perceptions of real depths at large observation distances. *Journal of Vision*, 10, 1–16.

Pirenne, M. H. (1970). *Optics, Painting, and Photography.* Cambridge, UK: Cambridge University Press.

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

Sedgwick, H. A. (2003). Relating direct and indirect perception of spatial layout. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), Looking into pictures: An interdisciplinary approach to pictorial space (pp. 61-75). Cambridge, MA, US: MIT Press.

Tscherning, M.H.E. (1904) *Physiologic optics: dioptrics of the eye, functions of the retina, ocular movements and binocular vision*. Philadelphia: Keystone Press

Vishwanath, D., Girshick, A. R., and Banks, M. S. (2005). Why Pictures Look Right When Viewed From the Wrong Place. *Nature Neuroscience, 8*, 1401–1410.

Vishwanath, D. (2013). Immersion, tangibility, realism: Explaining the qualitative experience of stereopsis. In: Human Vision and Electronic Imaging XVI, Rogowitz, B.E., Pappas, T.N. and de Ridder, H. (eds.). *SPIE Proceedings: Stereoscopic Displays and Applications XXIV*, 86480, 1-22

Wade, N. J., Ono, H., & Lillakas, L. (2001). Leonardo da Vinci's struggles with representations of reality. *Leonardo, 34*, 231–235.

Wandell, B. A., (1995) *Foundations of Vision*. Sinauer Associates, Sunderland, MA.

Wheatstone, C. (1838). On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision. *Philosophical Transactions of the Royal Society of London*, 128, 371-394.

Wijntjes, M., & Pont, S. (2012). Perceived depth in photographs: Humans perform close to veridical on a relative size task. *Journal of Vision, 12,* 277.

Footnotes

---

[i] Usage of the term “relative depth”. Two objects can be said to have the same relative 3D structure if they only differ by a uniform scaling factor. Two objects have the same affine 3D structure even if they differ by affine transformations (non-uniform scaling and/or shear)

[ii] In Koenderink et al. (1994) the authors tested themselves on a local slant estimation task (gauge figure task) comparing monocular, binocular and synoptic viewing of pictorial objects and found differences in perceived depth relief among the conditions. Vishwanath (2014) has discussed these results in light of the other more recent contrasting empirical results.

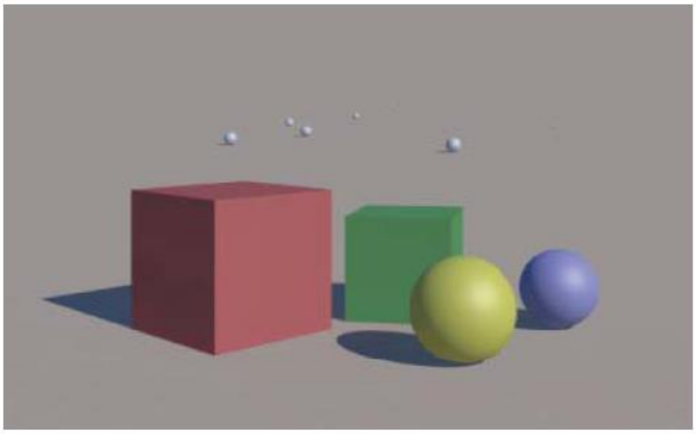

Figure 1.

A pictorial image of a 3-Dimensional scene

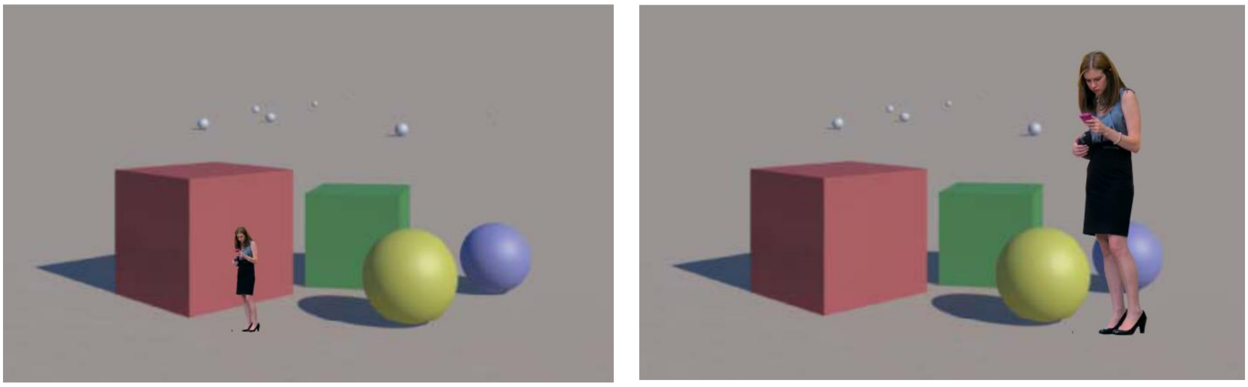

Figure 2

A perspective pictorial image of a 3-dimensional scene with the addition of a human figure as a familiar size cues to the scale of the scene. While the introduction of a familiar-sized object in a scene readily modifies cognitively inferred scale, no amount of cognitive effort alters perceived shape or layout of the depicted objects.