Toward a new theory of stereopsis

Dhanraj Vishwanath

School of Psychology & Neuroscience

University of St. Andrews

St. Andrews, Fife, UK

KY16 9JP

Email: dv10@st-andrews.ac.uk

Phone: 44 (0) 1334 46 2074

0. Abstract

Humans can obtain an unambiguous perception of depth and 3-dimensionality with one eye or when viewing a pictorial image of a 3-dimensional scene. However, the perception of depth when viewing a real scene with both eyes is qualitatively different: there is a vivid impression of tangible solid form and immersive negative space. This perceptual phenomenon, referred to as "stereopsis", has been among the central puzzles of perception since the time of da Vinci. After Wheatstone's invention of the stereoscope in 1838, stereopsis has conventionally been explained as a by-product of binocular vision or visual parallax. However, this explanation is challenged by the observation that the impression of stereopsis can be induced in single pictures under monocular viewing. Here I propose an alternative hypothesis that stereopsis is a qualitative visual experience related to the perception of egocentric spatial scale. Specifically, the primary phenomenal characteristic of stereopsis (the impression of 'real' separation in depth) is proposed to be linked to the precision with which egocentrically scaled depth (absolute depth) is derived. Since conscious awareness of this precision could help guide the planning of motor action, the hypothesis provides a functional account for the important phenomenal characteristic associated with stereopsis: the impression of interactability. By linking stereopsis to a generic perceptual attribute, rather than a specific cue, it provides a potentially more unified account of the variation of stereopsis in real scenes and pictures, and a basis for understanding why we can perceive depth in pictures despite conflicting visual signals.

1. Introduction

Depth and 3-D structure can be perceived when viewing a real scene with just one eye, or when viewing a picture, but they appear most qualitatively vivid when a real scene is viewed with both eyes. This qualitative difference has been described with great clarity by Susan Barry, a neuroscientist with infantile strabismus who recovered binocular vision in late adulthood. This is how she characterized her newfound visual experience (from Barry, 2009 & 2011):

> "[I saw] palpable volume[s] of empty space…I could see, not just infer, the volume of space between tree limbs…the sink faucet reached out toward me… the grape was rounder and more solid than any grape I had ever seen… Objects seemed more solid, vibrant, and real"

This visual experience is often called "stereopsis", a term which derives from the Greek words for 'solid' ("stereos") and 'sight' or 'appearance' ("opsis"). It refers to the characteristically vivid impression of solid form and immersive space that usually obtains under binocular viewing of real scenes. As Barry (2009) states,

> "When I first learnt about stereopsis in college, I wondered if I could imagine this way of seeing. Now I had my answer. I could not. Stereopsis provides a distinctive, subjective sensation…While I could infer indirectly a sense of depth through cues like perspective and shading, I could not synthesize stereoscopic depth [*sic*] from other visual attributes. The sensation provided by stereopsis of empty space and things projecting or receding into that space is unique."

Scientific interest in the phenomenon of stereopsis can be traced back to attempts to improve the depiction of depth in pictorial images during the Italian Renaissance. The development of perspective projection at this time made it possible to produce pictorial depictions that closely reproduced the visual stimulation generated when viewing a real 3-D scene (Kubovy, 1986; Pirenne, 1970; Wade, Ono & Lillakas, 2001). A major puzzle for Renaissance artists was why it was not possible to replicate the vivid qualitative impression of depth obtained in real scenes, despite accurate and realistic perspective rendering (Wade et al., 2001). In 1838, Wheatstone identified a link between this qualitative impression and the characteristic differences between the images of the two eyes when viewing real objects (binocular disparities). Based on this observation, Wheatstone invented the stereoscope, a device that allowed the pictorial simulation of binocular disparities by presenting two slightly different pictures to each eye (Wheatstone, 1838).

The difference between the impression of depth obtained in a single picture and that obtained with stereoscopic images can be appreciated in Figure 1. The remarkable vividness of depth achieved under stereoscopic viewing led to a longstanding belief that binocular disparities were the causal basis of the impression of stereopsis (Koenderink, 1998; Schlosberg, 1941), and the term *stereopsis* has come to simply imply depth perception from binocular disparities.

-Figure 1-

However, many researchers, including Wheatstone, had recognized that the impression of stereopsis can be obtained in the absence of binocular disparities, most particularly when a real scene is viewed with one eye while the observer moves (Helmholtz, 1866; Rogers & Graham, 1979; Wheatstone, 1838;). This observation has conventionally been attributed to the geometric similarity in visual stimulation between binocular viewing and monocular viewing under self-motion. In binocular vision,

different views of the object/scene are seen simultaneously with each eye (binocular parallax); during self-motion, different views of the object or scene are seen in temporal sequence (motion parallax). This suggests that it is the simultaneous or successive presentation of different perspectives of an object (visual parallax) that gives rise to stereopsis.

However, the parallax explanation is challenged by observations that the impression of stereopsis can also be induced under monocular viewing of a static pictorial image (Ames, 1925; da Vinci cited in Wade et al., 2001; Wheatstone, 1838). A persuasive example is viewing a pictorial image with one eye through a reduction aperture (Figure 2).

-Figure 2-

Surprisingly, only a handful of scientists in the last 150 years have described or examined this remarkable effect (e.g., Ames, 1925; Eaton, 1919; Judge, 1926; Michotte, 1948; Schlosberg, 1941; Smith & Smith, 1961; Wheatstone, 1838) and the phenomenon remains largely neglected in modern depth perception research (see Koenderink, 1998). There are two possible reasons for this. First, previous reports were based on the researchers' own introspections and no empirical validation of the effect in naïve subjects exists. Second, the *qualitative impression* of stereopsis is often conflated with the *quantitative magnitude* of perceived depth; a conflation that is often used to discount the significance of differences in qualitative depth impression among different viewing conditions (e.g., real scenes vs. pictures, monocular vs. binocular viewing).

Perceived *quantitative* depth is conventionally thought to be derived by combining estimates from individual depth cues in some statistically optimal manner (e.g., Buckley & Frisby, 1993; Bülthoff &

Mallot, 1988; Clark & Yuille, 1990; Landy, Maloney, Johnston & Young, 1995). The depth cues available from a single eye in static scenes are usually referred to as pictorial or monocular depth cues. They include interposition (occlusion), shading patterns, perspective scaling (relative size), perspective foreshortening, perspective convergence[1], texture gradients, and relative defocus blur. Binocular disparity is the primary depth cue obtained from combining information from both eyes.

The standard explanation for the difference in depth impression between real scenes and pictures is based on the relative coherence among depth values specified by monocular and binocular cues (Ames, 1925; Schlosberg, 1941). When a real scene is viewed binocularly, both types of depth cues coherently specify the same (or similar) depth values. However, under binocular viewing of a picture, the pattern of disparities specifies a flat surface which conflicts with the depth variation specified by the pictorial cues. According to the conventional explanation, the conflicting binocular information causes a reduction in magnitude, or 'flattening', of depth that would otherwise be perceived on the basis of the monocular cues (Ames, 1925; Young, Landy & Maloney, 1993), resulting in the lack of stereopsis (Ames, 1925). The induction of stereopsis under monocular viewing of a picture is ascribed to the elimination of the conflicting binocular information, resulting in a greater magnitude of perceived depth consistent with the monocular cues (Ames, 1925; Schlosberg, 1949; Koenderink, 1998; see also Young et al., 1993). This idea therefore links stereopsis to a greater coherence among depth cues and a greater magnitude of perceived depth.

In summary, there are three main hypotheses that have been put forth to explain the variations in perceived stereopsis in real scenes and pictures

---

[1] Perspective projection leads to two basic effects at the retinal image: foreshortening and scaling. However, it is customary to distinguish between different types of visual stimulation that may contain one of both of these effects and denote them as different cues (e.g. , convergence cue, texture cue, relative size cue)

1. *Binocular Disparity Hypothesis*: Stereopsis is the result of binocular vision and occurs

when binocular disparities specify the perceived depth relations (e.g., Ogle, 1950)

2. *Visual Parallax Hypothesis*: Stereopsis is the result of visual parallax: the simultaneous

or successive presentation of different views of the object/scene. Parallax can derive from

differences in the images of two eyes, or differences between images in a single eye over

time (e.g., Wheatstone, 1838; Rogers & Graham, 1979).

3. *Cue Coherence/Depth Magnitude Hypothesis*: The impression of stereopsis is directly

related to the magnitude of perceived depth: when cues coherently specify similar depth

values, a greater and more accurate magnitude of depth is perceived, resulting in a strong

impression of stereopsis. Conflict between depth cues results in a diminished magnitude

of perceived depth and the lack of stereopsis (e.g., Ames, 1925).

It is clear from the outset that there are significant problems associated with each of these hypotheses.

For example, the Binocular Disparity and Visual Parallax Hypotheses [the most widely held explanations

for stereopsis (Ponce & Born, 2008; Ogle, 1950; Rogers & Graham, 1979)] are unable to account for

previous reports of the induction of stereopsis in single pictures. The Cue Coherence/Depth Magnitude

Hypothesis conflates the fundamental distinction between qualitative and quantitative aspects of depth

perception central to the descriptions of the experience of stereopsis (e.g., Barry, 2009; Hibbard, 2008;

Vishwanath, 2011). It fails to explain why we clearly perceive substantial magnitudes of quantitative

depth in pictures despite cue conflict and the absence of the qualitative impression of stereopsis

(Vishwanath, 2011).

1.1 Stereopsis and the problem of terminology

The difficulties in explaining the causal basis of stereopsis is to a significant degree tied the lack of clear theoretical distinctions between *quantitative* and *qualitative* aspects of depth perception implicit in common terminology. While "stereopsis" arguably refers to a specific qualitative impression present under certain conditions, contemporary research typically makes little distinction between stereopsis as a qualitative perceptual attribute on one hand and binocular vision as a mode for deriving quantitative depth on the other. For example, here are three quotes drawn from a standard reference text in depth perception (Howard & Rogers, 2002)

"The term 'stereoscopic vision' [*sic*] means literally, 'solid sight' and refers to the visual perception of the 3-D structure of the world, when seen by *either one eye or two.*" [My emphasis]

"There are considerable similarities in the underlying computational theory of stereopsis based on disparity and that based on motion parallax."

"The term 'binocular stereopsis' is often used for the impression of depth arising from binocular cues. Unless otherwise specified, we use the term 'stereopsis' for depth impression created by binocular disparities."

Taken together, the broad interpretation of these quotes is consistent with the idea that stereopsis is not binocular vision per se, but a perceptual attribute that can derive from different sources of

information. However, it is unclear what the differences are among "visual perception of the 3-D world", "stereoscopic vision", "binocular stereopsis", and "stereopsis". For example, the first definition suggests that 'stereoscopic vision' can be achieved with one eye, and furthermore, that *any* condition in which there is perception of 3-D structure constitutes the perception of stereoscopic vision. But this interpretation misses the obvious qualitative difference between perceiving 3-D structure in single pictures and stereoscopic images ([Figure 1](#)), or, between viewing real scenes with one or two eyes (Barry, 2009). Moreover, while the first two quotes acknowledge that stereopsis can be caused by more than just binocular disparity, the third quote specifically links stereopsis to binocular disparity.

This ambiguity in terminology, widespread in the literature, suggests the lack of a clear theoretical framework for understanding the differences between qualitative and quantitative aspects of depth perception, monocular and binocular depth perception, real and pictorial depth perception. In this paper I differentiate between "visual perception of depth or 3-D structure", "stereoscopic vision", and "stereopsis". I define "visual perception of depth or 3-D structure" to generically refer to the perception of the 3rd dimension, in pictures or real scenes, with one eye or two, regardless of the cues inducing it or the qualitative impression associated with it. This includes the perception of both basic depth order and quantitative depth (for example, the perception of 3-D shapes and locations of objects in a 3-D pictorial space). "Stereoscopic vision" I take to literally mean 'vision through a stereoscope', or less literally, to refer to 'binocular depth perception': the visual perception of depth on the basis of binocular disparities. I reserve the term "stereopsis" for the following:

> The characteristically vivid impression of tangible solid form, immersive space and realness[2] that obtains under certain viewing and stimulus conditions.

---

[2] I distinguish here between 'realness', the sense that one is confronted with a real object or scene and 'realism', the attempt to visually depict things pictorially as they appear in the real world.

According to this definition, depth and 3-D structure can be perceived with one eye or two eyes, in pictures or real scenes, in the absence or presence of stereopsis. But stereopsis is a distinct qualitative impression only induced under certain conditions.

1.2 Stereopsis and the paradox of picture perception.

Since the Renaissance, the problem of understanding stereopsis has been intimately tied with the puzzle of pictorial depth perception (Wade et al., 2001). Pictures present a paradox. When we view a well-rendered picture or high-resolution photograph of a 3-D scene, we have a clear and unambiguous sense of the 3-D shape and layout of the objects in the scene. To a first approximation, the perception of 3-D structure in such images is crucially similar to that achieved viewing the actual scene. Significantly, there is no apparent distortion of 3-D object shape or layout in comparison to the real scene. The flatness of the picture surface may be apparent, but the objects in pictorial space do not themselves appear flat or squished. If they did, pictures and photographs would not serve the ubiquitous purpose they do. Moreover, there is extensive empirical evidence that pictures convey a consistent and measurable percept of quantitative depth and 3-D shape (Koenderink, van Doorn & Kappers, 1994; Koenderink, van Doorn & Wagemans, 2011; Kubovy, 1986; Vishwanath, Girshick & Banks, 2005). The crucial difference between pictures and real scenes appears to lie in the presence or absence of a qualitative attribute (stereopsis) regardless of the actual 3-D relations perceived. None of the hypotheses outlined above provide an explanation for this. The Binocular Disparity and Visual Parallax hypotheses suggest that "real" depth perception only occurs in the presence of binocular disparity or parallax, and that depth in pictures is perhaps a poor stepsister; deriving from some form of higher-order inference or habit (Wheatstone, 1838). The Cue coherence/Depth Magnitude hypothesis

fails to explain why we perceive fully articulated and unambiguous 3-D structure in pictures despite reliable depth signals specifying a flat surface.

1.3 An alternative theory of stereopsis

In this paper I will present an alternative theory for the source and functional role of stereopsis[3]. The theory is founded on three fundamental perceptual distinctions:

1. The perception of quantitative 3-D structure vs. the qualitative impression with which 3-D structure is perceived.
2. The perception of relative depth relations vs. the perception of absolute depth values.
3. The magnitude of perceived depth separation vs. the precision with which depth separation is perceptually estimated.

Consider Figure 3 in which the observer views a simple 3-D object (shown in profile) which consists of two rectangular planes connected with a dihedral angle θ and oriented at angle $\emptyset$. The physical depth dimensions of the object with respect to the observer can be defined by the values d_1 and d_2. Viewing the object produces a retinal image shown on the right in Figure 3.

-Figure 3-

A conventional view is that the goal of depth perception mechanisms is to derive the values θ, $\emptyset$, d_1 and d_2 (e.g., Landy et al., 1995). However, the depth cues in the retinal image(s) (e.g., shading,

---

[3] Aspects of this theory were initially sketched out in Vishwanath (2011).

perspective, disparity) cannot by themselves be used to derive the absolute values d_1 and d_2. Depth cues only specify relative depth relations: angles (θ) and depth ratios $(d_1:d_2)$[4]. For example, texture and perspective cues can uniquely specify the slant of a planar surface with respect to the observer[5] but not the actual size of the surface and absolute distances from one point on the surface to another. A very large surface at a far distance can produce the same perspective image as a smaller surface with the same slant at a near distance.

Therefore, on the basis of depth cues alone, the observer in Figure 3 can in principle perceive the 3-D shape or relative depth relations [angles (ϕ, θ) and depth ratios $(d_1:d_2)$] but not its absolute size and depth values d_1 and d_2 (Figure 4A). In order to perceive the values d_1 and d_2, the observer must have information about the egocentric distance of the object (D) to scale the relative depth information and size specified at the retina (Glennerster, Rogers & Bradshaw, 1996; Landy et al, 1995; Kaufman et al., 2006). If cues to egocentric distance are available, the 3-D shape, scale and absolute depth values can all, in principle, be derived (Figure 4B).

-Figure 4-

A further consideration is the precision with which absolute depth values can be perceptually estimated. This will depend on both the precision of the cues specifying relative depth (e.g., slant, curvature-in-depth) and the distance cues used to scale them. For example, the shading cue is thought to yield a more variable estimate of 3-D surface curvature than binocular disparity (Bülthoff & Mallot,

---

[4] Here, depth ratios are defined along the z-axis dimension without the small angle assumption (i.e., without assuming parallel visual rays). The depth ratios therefore restrict object shape to a unique 3D interpretation up to a scaling factor. When parallel rays are assumed, depth ratios only specify the 3D shape up to an affine depth stretch.
[5] On the basis of certain assumptions such as rectilinearity, parallelism and surface continuity.

1988; Vuong, Caudek & Domini, 2006). Similarly, cues to egocentric distance can be more or less precise. For example, distance information derived from sensing the focus state of the ocular lens (accommodation) is considered to be significantly less reliable that distance information derived from monitoring binocular convergence (see Howard & Rogers, 2002). The precision (statistical reliability) with which absolute depth values are estimated will therefore vary as a function of the reliability of the constituent depth and distance cues.

Here I propose that *3-D structure* can be perceived solely on the basis of relative depth estimates, but that the impression of *stereopsis* is induced only when absolute depth values can be estimated. Furthermore, I propose that the strength of the impression of stereopsis that is experienced will be a direct function of the precision (statistical reliability) of the estimates of absolute depth; independent of the relative depth structure perceived. In conditions where distance cues are unavailable or highly unreliable, only relative depth relations are perceived yielding no impression of stereopsis; imprecise estimates will yield a weak impression of stereopsis while high-precision estimates will yield a strong impression of stereopsis (Figure 5). I will henceforth refer to this explanation as the *Absolute Depth Scaling Hypothesis*

-Figure 5-

In a later section I will discuss how the overall impression of stereopsis appears to consist of several primary, secondary and tertiary qualitative visual attributes. In terms of the link between absolute depth precision and stereopsis, we focus specifically on the two primary visual attributes of stereopsis:

the vivid impression of solid object shape and the impression of a palpable invisible (negative) space between objects. These attributes can be understood as both arising from an impression of "real separation" between points in depth; either within an object/part (perception of solid shape) or between objects/parts (the impression of a palpable invisible space). Operationally, it is the strength of this impression of 'real separation in depth' that is postulated to vary as a function of the precision of absolute depth estimates. More concretely, the strength of the impression of real separation between two points is postulated to be inversely correlated with the coefficient of variation (CV) of the estimates of absolute depth between the points. The coefficient of variation (CV) is a normalized measure of the statistical variation of an estimate (the standard deviation divided by the expected value) and is often referred to as the Weber fraction. In the simplest form we can write:

$$\zeta \sim \frac{1}{c_v} \qquad (1)$$

Importantly, under this hypothesis, the difference in the impression of stereopsis between two viewing situations does not necessarily imply a difference in perceived 3-D shape or depth relief. Also, the perception of a larger absolute depth value does not necessarily imply a stronger impression of stereopsis: a large absolute depth value perceived imprecisely is expected to produce a weaker impression of stereopsis than a small depth value perceived very precisely (Figure 6)

-Figure 6-

Several features of this hypothesis can be identified that overcome drawbacks of previous explanations. First, since it links stereopsis to the generic derivation of depth estimates rather than to specific types of

depth cues (e.g., binocular disparity) it can explain the presence of stereopsis under a range of different viewing conditions and cues. Second, in linking stereopsis to *precision* rather than *magnitude,* the hypothesis can explain the dissociation between perceived depth magnitude and stereopsis. Third, the proposition that relative depth relations are perceived independently of absolute depth provides a potential explanation for why pictures can convey an unambiguous perception of 3-D shape and relative layout despite the absence of the impression of stereopsis. Finally, the theory links stereopsis to the representation of absolute (egocentric) spatial parameters which are crucial for efficient visually guided action. It therefore provides a functional basis for understanding the secondary qualitative attributes associated with stereopsis which appear related to the capacity for motor interaction: the sense of tangibility, immersion and reality.

1.4 Relation to previous work

The measurement of the precision of perceptual judgments has been central to developments in vision research, but there have been few suggestions in the literature that precision (or statistical reliability) can be qualitatively experienced (though see Vishwanath, 2005). Binocular vision is conventionally regarding as the condition which yields the most precise estimates of quantitative (metric) depth (Howard & Rogers, 2002), but a systematic link to qualitative aspects of depth perception is rarely made. Within the philosophy-of-perception literature there have been suggestions that binocular stereopsis appears related to a greater "determinacy" or "extra way" in which depth is perceived (Tye, 2002; Peacocke, 2002). Hibbard (2008) has drawn attention to the potential link between these observations and the notion of statistical reliability or precision within the depth perception literature, and furthermore, highlighted the need for a more grounded theoretical understanding of qualitative aspects

of depth perception[6]. Michotte drew attention to the difference between the generic perception of 3-D

structure and the specific impression of reality in stereopsis. He linked the latter to visual characteristics

related to motor interaction: "the potential for being manipulated" (Michotte, 1948). To date, however,

no specific theory has been proposed that seeks to provide a unifying account of the variation, causal

basis and functional role of stereopsis. Importantly, no previous study has systematically analyzed the

qualitative variation of depth impression in real scenes and pictures, and linked these variations to

conceptual distinctions between relative and absolute depth, depth magnitude and depth precision,

qualitative impressions and quantitative depth judgments.

The second aspect of the current theory, the postulated dissociation between the perception of

relative depth relations and absolute depth values, finds sympathy in prior work. There have been

several propositions that the perception of 3-D space consists of multiple distinct or dissociated

representations; related to stages of information processing, or tied to specific adaptive roles. For

example, some authors have suggested that depth processing proceeds in a constructive fashion, from

the initial detection of depth order, to the derivation of quantitative relative depth relations, to the final

derivation of absolute depth values (Glennerster, Rogers & Bradshaw, 1996). They proposed that

different visual functions may have access to different stages of this processing. Others have suggested

that there may be a fundamental dissociation in the processing of relative and absolute (egocentric)

aspects 3-D space/objects (e.g., Marr, 1982; Zimmerman, Legge & Cavanaugh, 1995; Koenderink & van

Doorn, 2003; Loomis, Philbeck & Zahorik, 2002; Svarverud, Gilson & Glennerster, 2012), which

potentially links to the distinction between the use of visual information for motor interaction vs. its use

for object recognition (Milner & Goodale, 2008; Ungerleider & Mishkin, 1982). Picture perception has

---

[6]Paul Hibbard and I became aware of our mutual interest in qualitative aspects of depth perception in April 2008 when we
exchanged manuscripts [Hibbard (2008; in press at the time) and Vishwanath (2011; under review at the time] and we
subsequently collaborated on a study (Vishwanath & Hibbard, 2013).

often been suggested to induce a "dual percept": the simultaneous perception of the flat picture surface and pictorial depth (see Hecht, Schwartz & Atherton, 2003; chapters 1-5). The current theory provides a potentially more systematic account of this phenomenology.

The aim of this paper will be to describe and analyze the variations in perceived stereopsis for real scenes and pictures to show that they are inconsistent with prior explanations of stereopsis, but can potentially be accounted for by the present theory. It will also describe how the present theory provides a more grounded explanation of the phenomenal characteristics associated with stereopsis, and a clearer foundation for understanding why we are able to perceive depth in pictures in the first place.

2. Stereopsis in Static Single Pictures

2.1 Stereopsis induced by monocular-aperture viewing of pictures

Previous reports of the induction of stereopsis in pictures were based on expert observer introspections and no empirical validation of the similarity between monocular and binocular stereopsis has previously existed. In a recent study, we aimed to redress this by empirically establishing the qualitative characteristics associated with stereopsis in naïve subjects and determining if these same characteristics are reported for monocular and binocular stereopsis (Vishwanath & Hibbard, 2013). Stereonormal subjects were first screened to ensure that they were able to perceive and report on qualitative aspects of binocular stereopsis by comparing binocular and monocular viewing of a real complex object at a near viewing distance (60cm). The screened subjects then provided verbal reports comparing binocular and monocular-aperture viewing of colour photographs of natural scenes (1.3cm aperture). The large majority of screened subjects (20/23) reported a better impression of stereopsis under monocular-aperture viewing [χ^2 (2, N=23)=28.125, p<0.0001]. Both verbal and written (Table 1) descriptions of the

effect of monocular aperture viewing were highly consistent with the qualitative characteristics

conventionally attributed to binocular stereopsis, and very similar to descriptions of binocular stereopsis

in recovered strabismics (compare Table 1 and quote on page 1). Verbal responses (see Table 1 in

Vishwanath & Hibbard, 2013) included descriptions of a clearer and more definitive sense of depth, a

clearer impression of depth separation, a sense of realness, a sense of protrusion towards the viewer,

the feeling that things appeared touchable, and a few direct references to binocular stereopsis.

Table 1 (From Vishwanath & Hibbard, 2013; Table 3) Verbatim Written Descriptions of the Visual Effect

of Monocular-Aperture Viewing of Color Photographs of Real Objects for a Sample of 3 Naïve Subjects

Observer	Description
Subject 1	"With one eye, it seems as if I can touch the raindrops. 3-D illusion."
	"With one eye, it seems like the leaves are coming out of the screen and pointing at me."
	"With two eyes, image very flat—definitely on a computer screen, whereas the image with one eye looks like someone placed a plant in front of me."
Subject 2	"The leaves appeared more lifelike, as though they were right there and could be touched."
	"The gaps between the leaves had more depth and appeared more hollow."
	"The leaves jumped out of the screen much more."
Subject 3	"When viewing the objects through the aperture, the objects seems more 3-D"
	"Through aperture, the definition between objects is clearer."
	"Space in general is more defined."
	"Depth is more defined to the eye."

We tested two additional groups of naive subjects on Likert-style questionnaires where they affirmed or

denied 14 descriptions of visual characteristics on a seven-point visual analogue scale (see Botvinick &

Cohen, 1998; Ehrrson, 2007). Seven of the descriptors were target statements describing visual

characteristics associated with stereopsis while the other seven served as controls for suggestibility and

task compliance. The first group compared binocular viewing of single photographs and stereoscopic

color anaglyphs viewed though red/green filters, while the second group compared binocular and

monocular-aperture viewing of single photographs. Results for both groups were very similar, with

subjects affirming the target descriptors with a high degree of statistical significance (Figure 7; from

Vishwanath & Hibbard, 2013). For control items, ratings were either significantly less than or

insignificantly different from 0 for both comparisons.

-Figure 7-

These results, along with additional controls (Vishwanath & Hibbard, 2013), provide clear empirical

evidence that the impression of stereopsis can be induced monocularly in the absence of binocular

disparity or motion parallax[7].

2.2 Conventional Explanations of Monocular Stereopsis

The standard explanation for monocular stereopsis (the Cue Coherence/Depth Magnitude

hypothesis) attributes it to the removal of the binocular cues that contradict and "flatten" the depth

specified by pictorial cues (Ames, 1925; Kubovy, 1986; Koenderink et al., 1994). We tested this

explanation by measuring the variation of both the qualitative impression of stereopsis and perceived

depth magnitude under different viewing conditions (Vishwanath & Hibbard, 2013).

---

[7] Overall, nearly 90% of a total of 60 naïve subjects tested in various experiments reported percepts consistent with stereopsis under monocular-aperture viewing of picture.

To test the claim that conflicting binocular cues are responsible for the lack of stereopsis, we had naïve subjects rate the differences in the strength of stereopsis for different pairs of four viewing conditions (monocular, binocular, monocular-aperture, and binocular-aperture) when viewing single pictures. In the binocular and monocular conditions, the frame of the display was visible. In the monocular-aperture and binocular-aperture conditions, observers viewed the display through oval apertures (1.2cm diameter) that eliminated the view of the rectangular image boundaries. Subjects numerically rated the difference in impression of stereopsis within pairs of viewing conditions (e.g., binocular versus monocular viewing) on a scale where '0' indicated no perceived difference and '5' indicated the difference perceived in the binocular versus monocular-aperture viewing comparison, designated as the reference comparison and viewed first for each tested image. According to the cue coherence/depth magnitude explanation, the largest difference ratings should be obtained in the binocular versus monocular viewing comparison and the binocular-aperture versus monocular-aperture viewing comparison because they entail the removal of the main conflicting binocular cues. In the monocular versus monocular-aperture viewing comparison and the binocular versus binocular-aperture viewing comparison, there is no change in binocular cues, so little or no difference in depth impression is expected[8].   However, the largest shift in depth impression was obtained when there was no change in binocular cue conflict in the (monocular versus monocular-aperture viewing) and the smallest shift was obtained in a condition where the conflicting binocular cues were removed (binocular versus monocular viewing; see Figure 8).

---

[8] The apertures eliminate the visibility of the image frame, but frame visibility is not a depth cue in the conventional cue-integration sense. Previous reports of the effects of the presence of a rectangular frame on depth perception (Eby & Braunstein, 1995; Reinhardt-Rutland, 1999) used impoverished and ambiguous stimuli which have regular 2D and 3D shape interpretations (a striped triangle or an outline rectangle) and which may have been subject to other effects such as shape capture. Therefore, they are difficult to interpret for complex 3d objects in the context of a standard understanding of depth cue integration. There is no study, to my knowledge, that has shown that a visible frame is a quantitative cue that modifies depth perception for complex 3D objects in pictures or real scenes.

-Figure 8-

In order to test the second claim linking the induction of stereopsis to a greater perceived magnitude of depth relief, we measured the perceived curvature-in-depth of elliptical hemi-cylinders defined by a pattern of randomly distributed dots (Young et al., 1993) under monocular-aperture and binocular viewing. The order of presentation of viewing conditions and stimulus parameters was randomized both between and within subjects. Contrary to the cue-conflict account, we found no main effect of viewing condition on perceived 3-D curvature. This was despite the fact that the task was sensitive enough to reveal a significant horizontal-vertical anisotropy in curvature judgment similar to previously reported anisotropies in the presence of disparity (e.g., Frisby, Buckley & Horsman, 1995) and despite all subjects reporting an enhancement of depth consistent with stereopsis under monocular-aperture viewing in post testing. This result shows that perceiving a stronger impression of stereopsis is not the same thing as perceiving a greater magnitude of depth or a different 3-D shape. Importantly, it suggests that conflicting binocular cues are largely ignored in the quantitative derivation of pictorial depth.

A related hypothesis that seeks to explain the induction of stereopsis in pictures--without making a necessary link to quantitative depth perception--is based on the notion of "picture surface awareness". According to this explanation, the presence of binocular disparity and frame visibility causes a heightened awareness of the mediating picture surface resulting in the disruption of pictorial depth perception (da Vinci cited in Wade et al., 2001; Eaton, 1919; Kubovy, 1986; Michotte, 1948; Pirenne, 1970). Monocular stereopsis is induced by reducing the factors responsible for the visibility, and hence awareness, of the picture surface. Therefore, this *Picture Surface Awareness Hypothesis* predicts that the largest shifts in depth impression should occur in comparisons that include the monocular aperture viewing condition which eliminates binocular disparity as well as frame visibility, which is what we found

(Figure 8). However this hypothesis does not explain why one aspects of depth perception (stereopsis) is affected by picture surface awareness, while another (perceived magnitude of relative depth) is not.

2.3 Explaining Monocular Stereopsis on the basis of absolute depth perception

An important but neglected aspect of the binocular viewing of pictorial images is that essentially no combination of retinal or extra retinal information can visually specify the absolute depth separation between objects in the pictorial scene. Why is this? When a picture is viewed binocularly, distance cues such as binocular convergence, vertical disparity and the accommodative state of the lens specify the distance of the visible picture surface (Ames, 1925; Watt, Akeley, Ernst & Banks, 2005), so there are no optical distance cues that specify the distance of pictorial objects (Vishwanath, 2011). Pictorial depth cues such as shading, perspective and interposition can specify the 3-D shape and relative layout of objects in the depicted scene, but without distance information, these cues cannot be scaled to derive absolute depth or size. For example, absolute depth values are entirely ambiguous in the image in Figure 9, despite the clear and unambiguous perception of 3-D depth relations.

-Figure 9-

Absolute depth judgements within pictorial space have to depend on the inference of distance based on familiar-size information. But evidence suggests that familiar size does not operate as an

independent quantitative visual distance cue, but only permits cognitive inferences of size and absolute

depth in pictures (Gogel, 1969b; Gogel & Da Silva, 1987; Predebon, 1993)[9].

The above arguments suggest that for binocular viewing of pictures, while 3-D object shapes can be

clearly perceived, their distance, scale and absolute depth values should remain optically unspecified

(Figure 10A), and according to the current hypothesis, should lack the impression of stereopsis.

Monocular-aperture viewing eliminates disparity driven vergence, but it does not eliminate residual

distance cues such as the accommodation response, which continues to be driven by the markings on

the picture surface[10]. Distance information may be available directly from accommodation (Gogel, 1961;

Fisher & Ciufredda, 1988) or indirectly via its effect on vergence (Kenyon, Ciuffreda & Stark, 1978). In the

absence of a visible picture surface, it is plausible that the brain attributes the accommodation response

to the pictorial objects, and assigns any associated distance information to them, allowing absolute

depth values to be derived (Figure 10B). It is likely that viewing conditions that yield only a partial

reduction in surface visibility (removing either binocular disparity or visible frame) will result in a greater

variability in the assignment of distance information to the pictorial objects and a correspondingly

reduced sense of stereopsis, as we found (Figure 8).[11]

-Figure 10-

---

[9] 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 & DaSilva, 1987; Predebon, 1993). Familiar size can bias manual reach distance in the presence of other cues to distance (McIntosh & Lashley, 2008) but the effect is only 20% of the size perturbation. This is similar to how figure-ground (an ordinal depth cue) can bias estimates of disparity derived depth (Burge, Peterson & Palmer, 2005), but it does not demonstrate independent use of the familiar size for distance perception.
[10] Not that the aperture is not a pinhole, so it does not, in principle, alter accommodative load.
[11] It also predicts that other viewing/display manipulation that reduce surface or frame visibility (decreasing pixellization by increasing resolution, increasing image size (while maintaining resolution), or increasing viewing distance while maintaining the visual angle) can contribute to an increase in the impression of stereopsis

In contrast to the *Picture Surface Awareness Hypothesis*, the current explanation provides a more plausible mechanistic link between surface visibility and the induction of stereopsis in pictures. Importantly, it is also consistent with two crucial aspects of our empirical results: the lack of a difference in perceived depth magnitude accompanying the induction of stereopsis and the change in perceived distance/size of pictorial objects.

2.4 Monocular stereopsis & changes in perceived distance and size

The apparent size of an object is thought to be determined by scaling the retina image by the estimated distance to the object (Holway & Boring, 1941; Kaufman & Rock, 1962; Gogel, 1963)[12]; for a given retinal image size, objects that are perceived as closer will appear smaller. Pictorial objects are usually cognitively inferred to be at some distance beyond the picture surface, consistent with the size of familiar object in the scene relative to the picture frame and the size of the picture itself. If distance information from accommodation (or from accommodation-driven vergence) is assigned to the pictorial objects when the picture surface is less visible, the objects should appear to be located at some distance close to the picture surface (Figure 10B). Since the objects continue to subtend the same retinal image size, observers should perceive objects as appearing closer and smaller than cognitively inferred under normal binocular viewing. The significant majority of observers who report stereopsis under monocular-aperture viewing also report a change in perceived distance and scale of objects (17/20; $\chi^2(1,N=20)=9.8$, p<0.01 ; Table 2) [13].

---

[12] There has been a long and contested debate regarding the relationships among retinal size, apparent size, and apparent distance, particularly because of the possible interactions among them in the reporting of distances and sizes (e.g. Epstein, Park & Casey, 1961; Kaufman & Rock, 1962; McCready, 1965; Gogel, 1990). I will not address this complex debate here as it is outside of the scope of this manuscript. While specific predictions may be affected, the general validity of the current hypothesis does not depend on the outcome of this debate

[13] A few subjects reported perceiving smaller but more remote objects, or closer but bigger objects. This pattern of responses is consistent with reporting biases linked to the size-distance paradox, (Epstein et al., 1961; Gogel, 1990; Mon Williams & Tresilian

Table 2 (from Vishwanath & Hibbard, 2013) Subjects' Verbal Descriptions of the Perceived Differences in Size or Distance Under Binocular and Monocular-Aperture Viewing of Photographic Images.

Difference & specific verbal description	Number of Subjects
Closer	15
Looks closer (8)	
Looks zoomed in (4)	
Came closer/ came closer towards you (2)	
Like telescope view (1)	
Smaller	11
Looks smaller (8)	
Shrank (1)	
Child-like/ like an architectural model (1)	
Like miniaturized photographic scenes using blur (1)	
Like a farther away real scene	3
Smaller & more distant	2
Closer & Bigger	2

Note: Descriptions are drawn from reports by 20 subjects who reported a better impression of depth in the monocular-aperture condition. Size and distance were treated as a single category because of known dependencies and interactions in reports of these two attributes (the *size-distance paradox*). Each description refers to the monocular-aperture condition unless otherwise indicated.

1999b). Observers often paradoxically report that objects appear more remote because they appear smaller; or larger because they appear closer. This reflects the complications of assessing verbal reports of perceived size and distance. We therefore queried observers on size and distance jointly, due to these known interactions in reporting the two visual attributes.

Moreover, some observers report an impression of miniaturization in some images consistent with

an accommodation-related micropsia. This effect can be readily appreciated in Figure 11.

-Figure 11-

While the current explanation ascribes pictorial object scaling to a reassignment of accommodation-

based distance information, it is likely that other factors that affect perceived distance are also at play.

These include two important phenomena related to distance perception: the equidistance tendency

(Gogel, 1969a; Gogel & Teitz, 1977) and the specific distance tendency (Gogel & Teitz, 1973). In the

equidistance tendency, objects farther away from, or closer to, the viewer appear displaced toward the

point of fixation or attention (Gogel & Teitz, 1977; Gogel, 1990). It is plausible that an equidistance-type

tendency might play a role under monocular aperture viewing due to the visible blurred boundary of the

aperture located close to the eye. Such an effect should result in an estimated distance that is

potentially closer than the picture surface distance (derived from accommodation). The specific distance

tendency refers to the finding that when cues to distance are absent or severely degraded the fixated

objects appears to be located, quite consistently, at a distance of about 2-3m (Gogel & Teitz, 1973;

Gogel, 1990). The equidistance and specific distance tendencies could potentially contribute a distance

estimate (albeit erroneous) when pictures are viewed monocularly through an aperture at distances

beyond the range where accommodation provides useful distance information, and contribute to the

induction of stereopsis under monocular viewing even when the picture is located at distance greater

than 2m. While the specific combination of mechanisms that yield monocular stereopsis likely entail a

complex interaction of several factors involved in distance perception, our empirical results show that, consistent with the current explanation, the induction of monocular stereopsis causes a changes in perceived distance and scale, but not shape. These changes, and lack thereof, have not been previously appreciated in the introspective reports of expert observers (e.g. Ames, 1925; Schlosberg, 1949).

2.5 Stereopsis & depth-of-field blur.

As discussed earlier, in typical pictures or photographs, there is no optical information that specifies distances to objects. However, recent evidence has established that the pattern of defocus due to depth-of-focus effects is a quantitative for egocentric distance perception (Held, Cooper, O'Brien & Banks, 2010; Vishwanath & Blaser, 2010). When an object is fixated, it is brought into sharp focus by the accommodation of the lens, but regions in depth towards and away from the observer (relative to the object) will be out of focus. Depth-of-focus (DoF) refers to the limited region in depth surrounding the fixated object for which defocus blur is not noticed. At near fixations, DoF is small, creating a noticeable pattern of blur away from the central region of fixation. DoF increases with fixation distance causing a corresponding diminishment of DoF blur around the point of fixation. When blur gradients mimicking the shallow DoF of the human eye at near fixation (henceforth, DoF blur gradients) are applied to pictorial images they cause pictorial objects to appear significantly closer and smaller (Held, Cooper & Banks, 2010; Vishwanath, 2007; Vishwanath & Blaser, 2010; see Supplementary Figure 2). According to the current hypothesis, this optical distance information from DoF blur could potentially scale pictorial depth yielding an impression of stereopsis.

To examine if DoF blur can induce and impression of stereopsis, we had a group of 12 naïve subjects compare perceived depth impression in black and white photographs of natural objects with or without

a vertical DoF blur gradient. The blur gradient was simulated for a planar surface slanted about the

horizontal axis (70°) viewed from 45cm (Vishwanath & Blaser, 2010) and was identical in all the blur

images. A generic vertical blur gradient can strongly affect perceived distance regardless of the actual

depth structure depicted in the image (Vishwanath & Blaser, 2010).   The Images (31x23cm) were viewed

binocularly at a distance of 60 cm from a chinrest with the display frame visible.

Subjects were first shown a sample image indicating a central region of interest (ROI) within which

the blur and no-blur images were identical in sharpness. They were instructed to compare perceived

depth in the ROI only, ignoring differences in visual sharpness outside the ROI. The subjects compared

the pair of images (blur and no-blur; binocular viewing) by viewing them back and forth 3 times, for 2

seconds per viewing. The order of presentation of the blur and no-blur versions was randomized across

5 different images tested, and the order of the images was randomized across subjects. For each image,

observers reported if they perceived a difference in the impression of *depth* between the first and

second image of a pair. If so, they were asked in which image the impression of depth was enhanced.

They then verbally described the perceived differences, viewing the images again as necessary.   The

significant majority of naïve observers (9/12; $\chi^2(2,12)=11.86$, p=0.003) reported obtaining a better

impression of depth in images with DoF blur. The descriptions provided by these 9 observers were

consistent with those obtained in monocular stereopsis and stereoscopic viewing (Table 3). Note that

this effect was obtained despite the fact that both images were viewed binocularly, and in the presence

of conflicting binocular cues.

Table 3 Subjects' Verbal Descriptions of the Change in Depth Impression (Binocular Viewing) for

Photographic Images with Simulated Depth-of-field Blur Compared with the Original Photograph without

Blur.

Difference and specific descriptions	Number of Subjects
Parts stick out	6
Leaves stick out towards me (1)	
Slightly comes out more from screen (1)	
Leaves in front stick out (1)	
Some parts seem to stick out (1)	
More sticking out of second plane (1)	
Flower comes further out of the picture (1)	
More depth separation/ deeper	5
Central flower further away than leaves behind (1)	
Feels deeper (1)	
Leaves more separated, not lying on top of each other (1)	
More depth between leaves and inside of plant (1)	
Seems more spatial, can feel something is behind the rock (1)	
Local features seem deeper	5
Indentations/pock marks seem deeper (2)	
Parts seem deeply inside the rock (1)	
Ridges more pronounced (1)	
Main central notch seems deeper (1)	
No-blur condition looks flatter	4
Looks uniform in depth, all the same (1)	
Looks flatter / flattish (2)	
Flattish like a painting, only see depth because…	
we know it should be there (1)	
More distinct separation	3
Better depth between parts in central region and background (1)	
More depth planes, planes more distinctly separated (1)	
Depth is more distinct/clear even though it appears…	
smaller in actual amount (1)	
Higher curvature	2
Leaves more curved, deeper (1)	
Curve in individual leaves more pronounced/ more formed (1)	
More realistic	2
Things are actually apart, depth really exists (1)	
Looks more realistic (1)	

Note: Data is for 9 naïve subjects who reported an enhancement in perceived depth in the images with depth-of-field blur (see text). Responses are grouped into descriptive categories consistent with reports in binocular and monocular stereopsis (Barry, 2009; Vishwanath & Hibbard, 2013). Numerals at right indicate the number of subjects who gave a report consistent with that category. Specific descriptions are indicated in the tabbed items. Most subjects provided descriptions in more than one category.

One possible explanation for this result is that blur merely acted as an additional relative depth cue. Blur gradients can increase perceived slant consistent with the direction of the blur gradient (Watt et al., 2005) and relative blur is considered a qualitative cue that enhances perceived relative depth ordering (Marshall et al., 1996; Mather, 1996). However, in the current images, the blur gradient was derived for a planar surface slanted about the horizontal axis and simply applied to all the images. Blur was therefore locally inconsistent with depth structure, and its direction often orthogonal to any salient global planar slant structure[14]. Based on standard cue-combination regimes, the inconsistent blur should have resulted in a decrease, not an increase, in perceived depth (see Frisby et al., 1995; Watt et al, 2005). Moreover, the descriptions of depth enhancement (Table 3) clearly indicate a qualitative shift consistent with stereopsis, and not a mere increase in relative depth magnitude.

Importantly, the effects of depth enhancement under monocular-aperture viewing and binocular viewing in the presence of blur gradients are very similar. In both cases the enhancement of depth appearance is accompanied by a reduction in perceived scale or distance. This can be appreciated by comparing the images in Figure 11. There is a similar qualitative shift in depth impression in the central region, as well as an apparent change of scale, for monocular aperture viewing of a no-blur image and binocular viewing of the DoF blur image.

-Figure 11-

---

[14] See Supplementary Figure S3

2.6 Other methods of inducing stereopsis in static pictures

Several viewing methods have been reported to generate an impression of stereopsis in static pictures. One widely reported technique is viewing a picture with a synopter, an optical device that uses prisms to deliver identical images of the picture to each eye resulting in a zero-disparity field [Ames, 1925; Claparede, 1904 (cited in Koenderink, 1988); Schlossberg, 1941; Koenderink et al., 1994]. Since there are no disparities or motion parallax signals specifying the perceived depth, it is inconsistent with both the *Binocular Disparity* and *Visual Parallax Hypothesis*. It is also inconsistent with the *Cue Coherence/Depth Magnitude Hypothesis*, since under a strict interpretation of cue-combination, the zero-disparity field generated under synoptic viewing is in conflict with the depth specified by pictorial depth cues[15] (see Frisby et al., 1995).

The effect is however amenable to the explanation put forth here. Binocular viewing of a picture generates a non-zero disparity field that specifies the shape of the flat picture surface. Synoptic viewing removes this non-zero disparity field and also delimits the field of view, thus reducing or eliminating picture surface visibility (Koenderink et al., 1994), as in monocular aperture viewing. Although viewing is binocular, the vergence signal is non-informative since synoptic viewing causes near-parallel orientation of the eyes (Koenderink et al., 1994)[16]. Thus, similar to monocular-aperture viewing, the induction of stereopsis can be plausibly ascribed to reassignment of distance information from accommodation and default distance tendencies to pictorial contents and an accompanying rescaling of pictorial space.

---

[15] A zero-disparity field is only produced in natural conditions when an observer is viewing a very distant scene (>500m; see Ogle, 1950) with no intervening objects. But objects viewed through a synopter appear located near the visual display.
[16] Parallel orientation of the eyes is consistent with distance from about 6m to infinity (Mon Williams & Tresilian, 1999a), and therefore provides no useful information about distance.

2.7 Monocular vs. binocular stereopsis

The current hypothesis suggests that the impression of monocular stereopsis in pictures should be weaker than binocular stereopsis (assuming similar viewing distances). Binocular viewing of real objects at near distances or stereoscopic viewing should yield precise estimates of absolute depth since relative depth separations are specified by a highly precise depth cue (disparity) which can be scaled on the basis of reliable estimates of distance derived from vergence and vertical disparity. Accommodation provides less reliable estimates of distance (Fisher & Cuifredda, 1979) and therefore scaling of pictorial depth on the basis of accommodation (or default distance tendencies) should yield less reliable estimates of absolute depth. Furthermore, the hypothesis suggests that monocular viewing of real objects at near distances should yield an impression of stereopsis comparable to that under monocular-aperture viewing of pictures, since distance information in these cases will also come primarily from accommodation.

Preliminary evidence generally supports the above predictions. In a previous study (Vishwanath & Hibbard, 2010) we had subjects rank order the strength of the impression of stereopsis in four viewing conditions [monocular-aperture and binocular viewing of pictures; monocular and binocular viewing of a real object at the same viewing distance]. Subjects viewed the stimulus from a chin/head rest. They ranked binocular viewing of real objects highest, followed by monocular viewing of pictures and monocular viewing of real objects, both of which yielded similar mean rankings (Figure 12a). To further confirm this finding, we tested a new group of subjects who compared viewing a real object (flowering plant) and a high-resolution photograph of the same object for the same four viewing condition. The real object and the photographic image subtended the same visual angle viewed from 60 cm on a chin/head

rest[17]. 10 screened subjects[18] first rank ordered and then rated the relative strength of the impression of stereopsis in each of the four conditions on a 5-point scale where the lowest ranked condition was assigned '0' and the highest ranked condition was assigned '5'. This was done to normalize responses due to the large inter-subject differences expected (e.g., Koenderink et al., 1995). The subjects' ratings broadly supports the prediction that monocular stereopsis for pictures and real objects should be similar at near viewing distances, but weaker than binocular stereopsis (Figure 12b).   Median ratings for monocular viewing of the real objects and monocular-aperture viewing of the picture were the same and means were similar (3.81 and 3.65 respectively). Binocular viewing of the real object was however rated significantly higher than these two conditions [Kruskal-Wallis Test (χ^2=7.8, df=2, p=0.02)] [19].

-Figure 12-

3. Stereopsis in Moving Pictures

While the previous examples of the induction of stereopsis in pictorial images was restricted to static images, video sequences of rotational object motion on a 2Ddisplay have also been widely reported to generate an impression of stereopsis. The kinetic-depth effect is a simple moving 2D line or random-dot pattern which can elicit a vivid percept of solid form (Wallach & O'Connell, 1953). A related effect is the

---

[17] The photographic image was displayed on a 24" LCD display (1920x1200 pixel resolution). The photo was taken under the same lighting condition from the chinrest on which subjects were tested. The display and the nearest element of the real object were set at the same distance (60cm).

[18] Subjects were first screened to ensure that they could perceive monocular stereopsis. 2 out of 12 subjects reported perceiving no difference between monocular-aperture and binocular viewing of pictures, or monocular and binocular viewing of the real object and were not tested any further.

[19] The difference in ratings between monocular and binocular stereopsis may not reflect the true magnitude of difference in depth impression because the stimulus comparisons includes a condition where no stereopsis is present (binocular viewing of pictures) which could have caused a compression at the higher end of the scale. Perceptual or cognitive ratings of relative magnitude or value between two stimuli can be compressively biased by the presence of a third stimulus that has a significantly different value.

stereokinetic phenomenon in which a 2-dimensional rotating pattern produces a vivid percept of a 3-D conical shape rotating in depth (Mussati, 1924)[20]. These modes of depth stimulation are not strictly equivalent to motion parallax, since motion parallax arises from a laterally translating observer viewing a rigid environment; the depth-motion relationship is therefore not geometrically equivalent to KDE except in special cases[21].

The KDE has been shown to generate a compelling impression of stereopsis (operationally defined as "realness" and "sense of 3-dimensionality") under restricted monocular viewing (e.g., Proffitt, Rock, Hecht, & Schubert, 1992). While KDE stimuli do generate a percept of a rotating solid object under binocular viewing, it is unclear if an impression of stereopsis is also induced. This suggests that there is a distinction between KDE-type stimulation as a cue that specifies relative depth relations (like other pictorial cues), and the conditions under which it results in the impression of stereopsis. The impression of stereopsis widely reported under monocular viewing of KDE stimuli can plausibly be explained in similar way to monocular aperture viewing: the scaling of the pictorial depth generated by the KDE stimulus via residual distance information (accommodation, distance tendencies) when picture surface visibility is reduced or eliminated.

---

[20] http://www.michaelbach.de/ot/mot_ske/index.html
[21] KDE and Motion Parallax are sometimes subsumed under the general term "structure from motion". However they are not geometrically equivalent as depth cues. 3D shape interpretation (relative depth) from KDE is distance invariant (like static cues such as perspective), while depth interpretation from motion parallax is a function of square of estimated distance (like binocular disparity). Moreover, while KDE does not require estimation of any extra-retinal component, derivation of depth from motion parallax does: estimation of observer velocity (Ono et al., 1986).

4. Summary: Stereopsis in Pictures

Stereopsis can be induced in static and moving pictorial images in the absence of binocular disparities for a wide range of viewing conditions. Specifically:

1. The induction of monocular stereopsis in pictures does not appear to be a necessary function of the reduction of binocular cue conflict (Figure 8).   Stereopsis appears to be induced even in the presence of binocular cue conflict (the DoF blur condition, Table 3).

2. Monocular stereopsis in pictures does not entail a necessary change in perceived 3-D shape or an increase in magnitude of perceived depth.

3. The induction of stereopsis appears to be linked to a reduction in surface visibility. However, stereopsis appears to be induced even when the picture surface is visible (DoF blur condition, Table 3).

4. All forms of induction of stereopsis in pictures indirectly or directly imply the role of alteration of distance information. There is a change in perceived distance and/or scale accompanying stereopsis (monocular-aperture viewing) and stereopsis appears to be induced by introducing optical cues to distance perception (DoF blur).

These observations are contradictory to the *Binocular Disparity and Visual Parallax Hypotheses*. The lack of correlation between the impression of stereopsis and the degree of binocular cue conflict or perceived depth magnitude is inconsistent with the *Cue Coherence/Depth Magnitude Hypothesis*. While the link between the strength of induced stereopsis and picture surface visibility is broadly consistent with the Picture Surface Awareness hypothesis, that hypothesis fails to explain other observations and findings: the widely acknowledged difference in impression of stereopsis between monocular and

binocular viewing of real objects where no picture surface is present in either condition (e.g., Ogle, 1950; Ponce & Born, 2008) and the induction of stereopsis via DoF blur which we found even when cues to surface visibility are present. Most importantly, the Picture Surface Awareness explanation does not provide a mechanistic account of why surface visibility matters.

The *Absolute Depth-scaling Hypothesis*, in contrast, can account for each of the observations outlined above (1-4) on the basis of the link between stereopsis and the precision of scaled (absolute) depth.   Since the hypothesis links stereopsis to a generic depth attribute, it does not depend on the presence of any specific depth cue, but only the availability of distance information to scale depth. Since it links stereopsis to precision rather than depth magnitude, it readily explains the fact that stereopsis is not always accompanied by an increase in depth relief. Moreover, it provides a plausible mechanistic account for why surface invisibility plays a role in the induction of stereopsis. Table 4 summarizes the consistency of the various observations with the 5 hypotheses.

Table 4. The Capacity for Five Different Hypotheses to Explain Variations in the Impression of Stereopsis when Viewing Pictures. .

PICTURES	Binocular Disparity Hypothesis	Visual Parallax Hypothesis	Cue-Coherence/ Depth-Magnitude Hypothesis	Picture Awareness Hypothesis	Absolute Depth Scaling Hypothesis
Lack of stereopsis under Binocular Viewing	•	•	•	•	•
Stereopsis induced under Monocular-Aperture viewing			•	•	•
Variation in ratings for the impression of stereopsis comparing different viewing conditions				•	•
Lack of a change in perceived 3-D shape accompanying induction of stereopsis				○	•
Change in perceived size /distance accompanying the induction of stereopsis					•
Stereopsis induced in the presence of DoF blur gradients under binocular viewing					•
Stereopsis induced in stereo kinetic effect (SKE) and KDE displays under monocular viewing	•	•		•	•
Stereopsis induced under synoptic viewing			○	•	○

Note: Filled circles represent conditions that can be explained by each respective hypothesis, and outline circle represent conditions that can be partially explained.

5. Stereopsis in Real Scenes with a static observer

Viewing a real scene (or a stereoscopic image) with both eyes is considered to be the paradigmatic condition that yields an impression of stereopsis (Ogle, 1950; Ponce & Born, 2008; Wheatstone, 1838). Monocular viewing of real scenes is conventionally believed to not produce an impression of stereopsis (Ogle, 1950; Ponce & Born, 2008). However, as we have seen, preliminary evidence suggests that observers do obtain an impression of stereopsis under monocular viewing of real objects, albeit weaker than under binocular viewing (Figure 12). Moreover, previous reports by expert observers, as well as informal observations, suggest that the impression of stereopsis under binocular viewing of real scenes diminishes with viewing distance (Hildebrand, 1907; Ogle, 1950; Tscherning, 1904; Wheatstone, 1838). This observation can be demonstrated by situating oneself in a real scene with a wide range of object distances and comparing depth impression with one or two eyes (Ogle, 1950; Tscherning, 1904; see Figure 13). The strongest impression of binocular stereopsis is most often observed within a few meters, in what is sometimes referred to as *personal space* (Cutting & Vishton, 1995). Regions of space beyond a few meters (*action space*) yield a diminished impression of stereopsis, and for distant regions (*vista space*) there is little or no perceptible impression of stereopsis (Hildebrand, 1907; Ogle, 1950; Tscherning, 1904).

-Figure 13-

5.1 Variation in perceived stereopsis when viewing a real scene with both eyes

In order to empirically confirm the previously reported diminishment of stereopsis with distance, we had naïve subjects rate the perceived difference in depth impression comparing binocular and monocular viewing of real scenes for different object distances. We initially screened subjects by having them verbally characterize differences in depth impression between viewing a real object under binocular and monocular viewing. Only those who described visual differences consistent with binocular stereopsis were tested. Depth impression was further operationalized by having the subjects complete a questionnaire that identified the key qualitative characteristics of stereopsis (see Figure 7). In the rating task itself, subjects were first asked to compare the difference in depth impression (Monocular vs. Binocular) when viewing a complex object (potted plant) at a distance of 60 cm. The magnitude of the perceived difference in depth impression in this reference condition was given an arbitrary value of 5 units (Vishwanath & Hibbard, 2013). The observers then numerically rated the difference in depth impression (monocular vs. binocular viewing) for other groups of objects at different distances and depth separations.

In the first experiment, observers viewed the same scene (a street with trees and buildings) and rated depth impression for objects in the scene ranging in distance from 5-150m (distance to near point), and where the object(s) of interest subtended approximately 6-7 degrees of visual angle (horizontally). In the second test, a different group of observers rated depth impression for different scenes at different distances (15m to 250m) which subtended a horizontal visual angle of 8-9 degrees. The 15 and 30m distances were done in the same general area, with the observer viewing the scene from ground level. For the 40m and 250m views, observers viewed the scene from a parapet at an elevated location chosen to provide views of distant objects (Figure 14).

-Figure 14-

In both experiments, the reduction in the strength of the impression of binocular stereopsis with distance was statistically significant [Kruskal-Wallis Test; Left graph: χ^2=25.9; df=3;p<0. 00001; Right graph χ^2=27.4; df=4; p<0.0001; df=3; p<0. 00001). This reduction was independent of the magnitude of the actual depth separation among objects in the region under consideration (see figure 14). There was some consistency between the ratings in the two tests. For example, the judgements for the 30m and 40m distances, which involved very different depth intervals and object shapes, yielded similar mean ratings in both tests. However, the ratings from the elevated parapet location at 250m are comparable to those at 100m. Judgements of distance are dependent on eye elevation with respect to the ground plane as well as the visibility of a continuous ground terrain (Philbeck & Loomis, 1997; Ooi & He, 2007). Therefore, the two stimulus conditions may not be strictly equivalent. The effect of viewing context on distance/depth perception will be addressed in a later section.

5.2 Stereopsis in real scenes and the magnitude of binocular disparities

Both Ogle and Tscherning attributed the diminishment of stereopsis with distance to the reduction in the size of binocular disparities with increasing viewing distance (*Binocular Disparity Hypothesis*). Figure 15a plots the size of disparities as a function of distance for different fixation distances, based on a standard formulation of the geometric relation between fixation distance (Z_0), interocular distance (I), signed depth separation (z) from point of fixation, and disparity (δ) in radians (small-angle approximation).

$$\delta \approx I \left(\frac{1}{Z_0} - \frac{1}{z + Z_0} \right) \qquad (2)$$

The plots shows that as the distance of fixation increases, the size of disparity for a give depth separation reduces markedly, consistent with the inverse-square relation between disparity magnitude and distance implicit in Equation 2. Beyond a hundred meters, even very large absolute depth separations produce uncrossed disparities that are very small (Ogle, 1950; Palmisano et al., 2010). Ogle suggested that stereopsis should no longer be apparent beyond 300-650m where the size of binocular disparities is below detection thresholds (Ogle, 1950).   However, while the *Binocular Disparity Hypothesis* can plausibly explain the diminution of stereopsis with distance, it cannot explain why monocular viewing of objects at near distances also appears to induce an impression of stereopsis (Figure 12).

-Figure 15-

The observed weakening of the impression of stereopsis with distance under binocular viewing is clearly inconsistent with the cue coherence/depth magnitude hypothesis, since depth impression appears to diminish despite the significantly larger depth separations specified by a given disparity value. Even small disparity values can specify very large absolute depth separations at farther distances. For example, at a fixation distance of 10 meters, a disparity of 10 arcminutes specifies a depth separation of 10 meters, while the same disparity value at 50 cm specifies only 1.5 cm separation (Figure 15b).

In summary, existing theories of stereopsis cannot account fully for the variation of stereopsis in real scenes under binocular viewing, and the reported (albeit weaker) impression of stereopsis under monocular viewing of real scenes (Figure 12).   We now examine how the alternative theory linking stereopsis to the precision of estimates of absolute (scaled) depth can better account for both these observations.

5.3 Precision of distance estimates and depth from disparity

The reordered version of equation 2 shown below (equation 3) implies that the precision of estimates of

absolute depth will depend on both the precision of the disparity signal itself as well as the precision of

available distance information.

$$z \approx \frac{\delta Z_0{}^2}{I - \delta Z_0} \qquad (3)$$

The precision of the disparity signal for most viewing conditions can be inferred from existing data on

detection and discrimination thresholds. The disparity detection threshold (i.e., discrimination of

disparities from a base disparity of zero) is of the order of hyperacuities (less than 10 arcseconds).

However, discrimination thresholds increase markedly for increasing base disparity (Ogle, 1950;

Blakemore, 1970). Therefore, the precision of disparity as an absolute depth cue varies as a function of

the absolute size of the disparity. Depth discrimination capacity breaks down beyond disparity values of

about 1.5 degrees which fall outside fusional limits (Blakemore, 1970). While qualitative judgements of

depth order can still be made beyond this point, depth magnitude can no longer be estimated

(Blakemore, 1970; Ogle, 1950)[22]. This implies that stereopsis (as construed in this paper) should not be

induced for disparities beyond these values (1.5 deg).

Determining the precision of distance estimation is more complex, given the variety of cues available

and the existence of complex contextual effects on distance perception (Gogel, 1990; Allison et al.,

2009). It is generally acknowledged that most direct cues to distance are quite limited in their range.

---

[22] Ogle referred to this capacity as "qualitative stereopsis". He used the term "qualitative" to distinguish the derivation of depth order from un-fused disparities to the derivation of quantitative depth from fused disparities. The latter, he termed "patent stereopsis". His use of the term "qualitative" in reference to stereopsis should therefore be distinguished from its use here, where it implies the qualitative impression of depth associated with patent stereopsis.

Accommodation, for example, is thought to be effective only within a meter or less and to provide an imprecise estimate of distance (Fisher & Ciufredda, 1979). Vergence could in principle be used to estimate distances up to 6m, but empirical evidence has only demonstrated its effectiveness in direct judgements of distance up to about 2m, with distances beyond 1m showing significant underestimation (Gogel 1961; Gogel and Tietz 1973; Foley 1977; Mon Williams & Tresilian, 1999a; Viguer, Clement & Trotter, 2001). However, in the presence of a visible ground surface, distances up to 20 meters can be judged with a surprising degree of accuracy and precision even using tasks that measure perceived distance indirectly (blind walking, see Loomis & Knapp, 2003 for a review). One source of distance information in this situation is the declination of the object from sensed eye level above the ground, sometimes referred to as the declination from the visual horizon (Epstein, 1966; Ooi, Wu & He, 2001; Philbeck & Loomis, 1997; Sedgwick, 1986; Wallach & O'Leary, 1982).   It is also plausible that when there are intervening foreground objects, estimates of relative distances (absolute depth) between objects derived from scaled disparities in near regions (where distance to objects can be precisely estimated) could be successively integrated to estimate distances to more distant object. This would be useful for rich scenes with visible ground textures and foreground objects, and for regions beyond the range where the direct distance cues such as convergence can be used to scale disparities (Allison, Gillam & Veccelio, 2009). These various mechanisms for distance estimation could potentially operate (regardless of accuracy or precision) to distances of 100 meters or more where disparity is still a useful signal (Ogle, 1950; Palmisano et al., 2010). Also, for the present theory, it is only the reliability of the distance estimate, independent of the accuracy of computed distance that is relevant for the degree of stereopsis perceived; whether or not conditions yields disparity-based depth constancy (Foley, 1980), is not expected to affect the strength of stereopsis as construed here.

There is limited empirical data that has explicitly measured egocentric distance discrimination thresholds for a range of distances (from personal to vista space). A popular compilation of depth estimation for different viewing distances was restricted to judgments of relative depth ordering (e.g., Cutting & Vishton, 1995). Kaufman et al. (2006) derived distance discrimination thresholds via both binocular depth and size discrimination tasks for distances from 2.5-20m. Distance judgement thresholds were inferred by deriving the change in distance required to just notice a difference in disparity specified depth, or a difference in size between two luminous targets. From these results they inferred that distance discrimination thresholds increase as a power function with an exponent of 1.8. This translates to a nearly linear increase in CVs from 3% to 12% for distances from 2.5 to 20m. However, other data has shown near-veridical distance judgements (bling walking) to 12m with little change in standard deviations (averaged over 7 observers) for these distances, which translates to an effective reduction in CVs with distance (Loomis et al., 1992).

5.4 Estimating the precision of absolute depth from disparity for different viewing distances

In order to calculate the coefficient of variation of absolute depth estimates from disparity as a function of distance, we will use known disparity discrimination thresholds and assumed distance estimation thresholds. Disparity discrimination thresholds increase from a value of about 10 arcseconds (for base disparities near zero) to over 10 arcminutes (for base disparities of 1 arc degree). Only the disparity pedestal range of 30 arcseconds to 90 arcminutes is considered, because disparities much greater than 90 arcminutes can no longer support quantitative depth perception (Ogle, 1950; Blakemore, 1970). Figure 15b plots absolute depth values as a function of distance for 9 different values of disparity ranging from 20 arcseconds to 90 arcminutes.

Here we are interested in depth perception for distances from .5m to beyond 100m, and since no data exists on distance discrimination thresholds for this range, two threshold functions are assumed: (1) CVs for distance estimation that increase linearly with distance from a value of 2% at 50cm to 15% at 200m (see Kaufman et al., 2006) (2) constant CVs of 5% for all distances.   The former (1) corresponds to the capacity to discriminate differences in distance of 1cm at 50cm viewing and 30m at 200m viewing. The coefficient of variation of absolute depth estimates as a function of disparity and viewing distance was derived using Equation 3 by sampling values of disparity and egocentric distance assuming Gaussian distributions (each with standard deviation values based on the assumed perceptual discrimination thresholds above)[23]. Figure 16 shows the estimated coefficient of variation of absolute depth estimates (based on 10K samples) plotted as a function of distance for the same 9 disparity values as in Figure 15b, simulated under the two different assumed distance discrimination thresholds.

-Figure 16-

According to the estimates shown in Figure 16, the CV of absolute depth estimates increases with viewing distance even for a fixed disparity value. Also, CVs increase sharply beyond a certain fixation distance for a given (achievable) disparity value, such that the CV for certain value of disparity (e.g., 30 arcminutes) exceeds the CV for a lower disparity value (e.g., 10 arcminutes) beyond a certain fixation distance (2 m).   Note that the ordinate is plotted in linear units while the abscissa is plotted in log units, which magnifies the steepness of the upturn of each curve.   Based on these estimated values, the current hypothesis predicts the following: (1) the impression of stereopsis will diminish with fixation

---

[23] We focus here only on judgments beyond fixation (uncrossed disparities), though a similar analysis could be done for crossed disparities.

distance *even* for a fixed disparity value; (2) assuming there is a maximum CV beyond which no stereopsis is perceived, then for any given disparity value, there is a fixation distance beyond which no stereopsis can be perceived[24]. Furthermore, the hypothesis make the counterintuitive prediction that for a given fixation distance, a larger disparity value (associated with a larger absolute depth separation) can sometimes yield a weaker impression of stereopsis than a smaller disparity value. To understand the third prediction better, assume an observer fixating an object at a particular distance, and comparing the perceived sense of qualitative separation (characteristic of stereopsis) between the fixated object and each of two other objects at different distances beyond fixation[25]. Under certain fixation distance and disparity combinations, the impression of qualitative separation between the fixation target and the closer object should appear to be stronger than that between the fixation target and the farther object, despite the larger absolute disparity value in the latter.

The current hypothesis therefore makes specific testable predictions about the variations in the impression of stereopsis in real scenes under binocular viewing. These predictions will need to be necessarily tested in real scenes, or with synthetic stimuli that accurately replicate the natural variation in both depth and distance cues with viewing distance. It therefore poses some difficult but not insurmountable technical challenges.

5.5 Stereopsis in real scenes under monocular viewing

Our preliminary results suggest that the impression of stereopsis can be induced at near viewing distances for monocular viewing of real objects (Figure 12). Monocular viewing eliminates two important distance cues (vergence and vertical disparity) as well as eliminating the precision depth cue

---

[24] This would correspond to the point beyond which only depth order can be conceivably be discriminated on the basis of disparities, but without the impression of stereopsis (what Ogle(1950) referred to as "qualitative stereopsis").
[25] This could be done, for example, by comparing the difference in perceived qualitative impression between viewing with one and two eyes.

of horizontal disparity. At very near viewing it may be possible to scale depth via distance information

from accommodation or accommodative vergence, as well as default distance tendencies, but such

estimates will likely be less precise than those achieved from disparity scaled with vergence or vertical

disparities. Monocular viewing with continuous ground plane information can provide reliable estimates

of distance to at least 7m (Wu et al., 2004), which suggests that depth could, in principle, be scalable

under monocular viewing well into action space. However, the depth cues specifying the relative

separation between objects are themselves considerably less reliable than disparity. Only texture

perspective has previously been shown to provide relative depth estimates with precision comparable to

disparity (Hillis et al., 2004). But such precision can only be achieved at higher slants and conditions

where the retinal extents of objects are large, i.e., near viewing or large slanted planar surfaces such as

the ground plane (see, Saunders & Backus, 2006). For individual objects or surface patches, the reliability

of this cue will decrease with distance simply because the retinal extent and perspective convergence

and foreshortening information both reduce with distance (Figure 17). Other sources of relative depth

information such as shading, relative size and interposition will either provide only ordinal information

(interposition) or are considerably less precise than disparity (e.g., shading). Monocular viewing of real

scenes should therefore yield significantly less reliable estimates of absolute depth, and a considerably

weaker impression of stereopsis than binocular viewing at distances beyond interaction space.

-Figure 17-

The degree of stereopsis induced under monocular viewing will no doubt be a complex function of

the available monocular depth cues and distance estimates. Default distance tendencies, changes in eye

height relative to the ground plane which alters declination information (Loomis & Philbeck, 1999;

Philbeck & Loomis, 1997) and visibility of the ground plane (Ooi and He, 2007) may further modulate

depth scaling.   In situations where the observer is viewing distant objects where declination and ground plane information is disrupted or unavailable, the visual system may erroneously scale relative depth cues on the basis of estimates deriving from the specific distance tendency. This is the tendency to ascribe a default distance of about 2-3 meters for objects in the absence of any other available distance information[26].   While the reliability (precision) of distance estimates deriving from the specific distance tendency has not been explicitly measured, it is conceivable that under certain conditions it may yield more precise (but erroneous) estimates of distance to very distant objects than would be yielded under normal viewing conditions. It could therefore induce a stronger impression of stereopsis than otherwise expected for a monocular observer located on the ground with visible terrain (Figure 19).

Since the specific distance tendency is thought to also occur for binocular viewing, particularly at far viewing distances beyond the range of vergence (Gogel & Teitz, 1973), similar counterintuitive increments in perceived stereopsis may occur also occur under binocular viewing in special viewing contexts such as in Figure 18. In such situation, the small disparity values, which are normally scaled by the relatively imprecise (but correct) estimate of distance, could be scaled by the relatively more precise (but incorrect) default distance estimate.

-Figure 18-

These complex effects of default distance tendencies may underlie the pattern of results we found comparing monocular and binocular viewing of real scenes at different distances (Figure 14), where elevated (parapet) viewing with an interrupted ground plane yielded higher depth impression difference

---

[26] Informal evidence for object scaling based on the specific distance tendency is the common observation that objects viewed from a known high vantage point (e.g., the window of a landing plane) often appear miniaturized.

ratings (binocular vs. monocular) than might have been expected from the trends in the condition with a visible ground plane.

6. Stereopsis in Real Scenes with a Moving Observer

Viewing a real scene with one eye under self-motion has long been acknowledged to produce an impression of stereopsis. In fact, simulated motion parallax displays where head movements are yoked to the image motion of random-dot corrugated surfaces can elicit depth impression that is as compelling as obtained for random-dot stereograms (Rogers & Graham, 1979). This observation has conventionally been attributed to the geometric similarity between the successive visual stimulation in a single eye under observer motion and the differences between images of the two eyes under static binocular viewing. But how might one explain the induction of stereopsis from motion parallax on the basis of the alternative hypothesis described here?

The most important similarity between disparity and motion parallax as depth cues is that they both necessarily require an independent estimate of egocentric distance in order to produce a unique estimate of depth[27]. In the absence of distance information, disparity and motion parallax do not specify a unique 3-D shape, but only specify an infinite class of shapes related by an affine stretch (Figure 19). In contrast, quantitative monocular depth cues can specify unique relative depth values (3-D shape) under appropriate assumptions, even in the absence of distance information[28]. The relationship

---

[27] Binocular disparities can, under certain conditions, be scaled by signal present in the retinal disparities, namely the gradient of vertical disparities (Rogers & Bradshaw, 1993).

[28] Some researchers have suggested that even monocular cues such as shading and texture only specify depth up to an affine transform (e.g., Lappin & Craft, 2000; Koenderink et al., 2012). Here we assume the conventional view that such cues can specify unique 3D shape as long as suitable assumptions are made (e.g., rectilinearity, Lambertian reflectance, etc.).

between motion parallax magnitude, distance and absolute depth is essentially the same as that for

disparity with the translational extent replacing interocular distance (Ono, Rivest & Ono, 1986)

-Figure 19-

Since motion parallax (as a relative depth cue) has comparable precision to binocular disparity (Rogers &

Graham, 1982) the presence of this cue should induce a strong impression of stereopsis as long as

reliable distance information is available to scale it (i.e., at near viewing). The main difference between

motion parallax and disparity is that while achievable disparity values (at a given distance) are

necessarily constrained by the fixed interocular distance (Equation 1), motion parallax magnitudes are

not; for a given depth structure and temporal interval, they increase with the velocity of observer. Large

values of parallax can be achieved, even at far distances where disparity values will necessarily be small.

Another difference is that while the presence of disparities cannot be attributed to an extrinsic physical

cause other than a difference in depth, the presence of motion parallax can be plausibly attributed to

non-rigid relative motion between objects. Unresolvable (large) disparities lead to diplopia or double

vision, however ecologically invalid levels of motion parallax (e.g. viewing from a fast-moving vehicle)

lead to the physically plausible interpretation of non-rigid relative motion between external objects, if

distance information is unreliable or parallax is quite large (Gogel, 1990; Ono et al., 1986) [29]. This

change from the impression of parallax-based stereopsis to relative motion should occur at much nearer

viewing distances under monocular viewing, since disparity and vergence are not available to

supplement distance estimation. Thus, the fact that motion parallax can induce stereopsis comparable

---

[29] This can be experienced informally when viewing a real scene from a fast-moving vehicle. Under monocular viewing, for locations beyond a few meters, there is the perception of non-rigid relative motion between objects at different depths instead of, or alongside, an impression of stereopsis.

to binocular disparity under certain conditions, but is diminished in others is consistent with the hypothesis proposed here.

7. Summary: Stereopsis in Real Scenes

While certain aspects of the impression of stereopsis in real scenes can be partially explained on the basis of previous hypothesis, the *Absolute Depth Scaling Hypothesis* provides a potentially more unified understanding of the observed variations in the impression of stereopsis in real scenes for both static and moving observers, as well as under binocular or monocular viewing.   Moreover, the hypothesis provides further testable predictions (Figure 16), and can accommodate counterintuitive observations that arise when viewing context (e.g. changes in viewing elevation, visibility of a continuous ground plane) is altered. Table 5 summarizes the consistency of the 5 different hypotheses with the observed variation in stereopsis in real scenes.

Table 5. The Capacity for Five Different Hypotheses to Explain Variations in the Impression of

Stereopsis in Real Scenes.

REAL SCENES	Binocular Disparity Hypothesis	Visual Parallax Hypothesis	Cue-Conflict/ Depth-Magnitude Hypothesis	Picture Awareness Hypothesis	Absolute Depth Scaling Hypothesis
Difference in the impression of stereopsis between monocular and binocular viewing	●	●	●		●
Impression of stereopsis under monocular viewing			●	○	●
Variation in the impression of stereopsis with viewing distance (monocular viewing)			○		●
Variation in the impression of stereopsis with viewing distance (binocular viewing)	●	○			●
Stereopsis induced by motion parallax		●	●		●
Variation in the impression of motion-parallax based stereopsis with availability of distance information					●

Note: Filled circled represents conditions that can be explained by the respective theory. Outline circle

represent conditions that can be partially explained.

8. Stereopsis, Perceived Depth Magnitude & Depth Phenomenology

.

8.1 Seeing depth in the absence of stereopsis

Beyond the underlying cause of stereopsis, a critical question is why we clearly and unambiguously perceive metric relative depth structure (3-D object shape and layout) in conditions that fail to induce stereopsis (e.g. binocular viewing of a picture; viewing a real scene with one eye beyond interaction space). What is the nature of the depth percept in the absence of stereopsis? All previous theories of stereopsis and quantitative depth perception are essentially silent on this matter. They are limited to specifying the conditions in which stereopsis can be induced but make no predictions regarding the nature of perceived 3-D structure in the presence and absence of stereopsis.

For binocular viewing of a picture, the Cue Coherence/Depth Magnitude hypothesis (understood in the context of current models of depth-cue integration) predicts the perception of flat patterns on a planar surface, or at best, the perception of flattened 3-D object shapes (Young et al., 1993; Landy et al., 1995; Bülthoff & Mallot, 1988). This is because such theories predict that perceived depth should be a statistically optimal or robust interpretation given the available depth signals: the most reliable set of visual signals, or the ones that jointly specify the same structure, should determine perceived 3-D form. When a picture such as Figure 20 is viewed from a relatively close distance (say, 50 cm) almost all available depth signals (disparity, motion parallax due to small head movements, pixellization, vergence changes with successive fixations and focus cues) specify a flat planar surface. Only one conventional cue[30] (shading) specifies a 3-D shape in the image shown in Figure 20, but it is a cue that is generally regarded to be significantly less reliable than binocular disparity (Bülthoff & Mallot, 1988; Vuoung,

---

[30] Computational assumptions used to interpret depth cues, such as rectangularity, orthogonality of surfaces, or parallelism of lines, do not constitute depth cues as conventionally understood in depth cue integration models.

Caudek and Domini, 2006).   Moreover, recent statistical models of depth cue integration posit the

existence of a "flatness prior" which claims that the visual system is biased towards inferring flat

frontoparallel surfaces in the presence of unreliable or conflicting depth cues (Adams & Mammasian,

2004; Domini, Shah & Caudek, 2011; Watt et al., 2005; Young et al, 1993).   Since the majority of cues,

including the most reliable one (disparity), specify a flat surface, it would be difficult for cue integration

models to adequately explain why we perceive a 3-D object instead of a flat patterned surface.   While an

appeal could be made to a "dual percept" state where disparity and pictorial information are

experienced separately, such a scenario is not consistent with the claims of the cue-combination

literature which concluded that depth cues (e.g., disparity & texture) are mandatorily combined when

arising from a single modality such as vision (Hillis, Ernst, Landy & Banks, 2002).

-Figure 20-

8.2 Do we perceive "flattened" depth in pictures?

The alternative possibility is that statistically optimal cue integration yields the percept of "flattened"

3-D structure when pictures are viewed binocularly. However this idea is challenged by our normal

experience with pictures. We routinely see pictures of objects we have never seen before and in the

subsequent encounter with the real object, we seldom notice significant differences in perceived 3-D

shape: we do not have the impression that objects or spaces in pictures are considerably "flatter" or

more "sheared" than they are in real life[31]. Since we are quite sensitive to distortions in pictorial object

shape (~5% change of an object dimension; Vishwanath et al., 2005), any significant depth compressions

or shape distortions under binocular viewing of pictures should be easily discernible. If depth in pictures

was perceptually flattened, the pictorial depiction that most closely reveals the correct shape of an

object should have an artificially exaggerated z-component. It predicts that direct photographs of

objects will not provide the best pictorial rendition of true 3-D object shape, but instead, that the

photographer or artist should purposely exaggerate the z-dimension by using projective or pictorial

tricks. In fact, the truest depiction of object shape is thought to be a photograph taken with a 50-60mm

focal length lens (Pirenne, 1970; Kubovy, 1986; Vishwanath et al., 2005). Moreover, we can all agree on

whether a pictorial depiction of a 3-D object matches the real 3-D object—precisely what separates good

drawings from not so good ones. Such judgments do not seem to depend on whether the depiction is

viewed with one eye or both, or whether stereopsis is present or not[32]. This lack of difference in

perceived global 3-D shape, between monocular and binocular viewing, is independent of familiarity,

see, for example, [Figure 20].

   The most recent evidence that has measured global judgments of depth or 3-D shape in pictures

corroborates these anecdotal observations. Quantitative judgements of curvature-in-depth (Vishwanath

& Hibbard, 2013), pictorial depth range (Wintjes & Pont, 2012), dihedral angle between planar surfaces

(Cooper & Banks, 2012) and slant of planar surfaces (Erkelens, 2012) show little or no difference

between binocular, monocular, and monocular aperture viewing.

---

[31] Here we assume a picture with a "normal" projection and excluding situations where camera parameters themselves cause projective distortion (wide angle distortion, keystone effects, anamorphic effects; see Kubovy, 1986; Pirenne, 1970; Vishwanath et al., 2005)

[32] An artist sometimes closes one eye when copying a real scene onto a picture, yet in the final picture the depicted objects appear to accurately match the 3D shape of the real object, even when viewed with both eyes. This lends more support to the idea that perceiving 3D surface shape is not linked to stereopsis. Stereopsis could however interfere with the transfer of shape outlines from percept to paper, and explain why copying from pictures or with one eye when viewing real scenes is much easier.

One body of data, however, does appear to support the view that conditions which induce little or

no stereopsis yield a flattened perception of shape (Doorschot, Kappers & Koenderink, 2001; Koenderink

et al., 1992, 1994, 1995; Koenderink, van Doorn, Kappers & Todd, 2001). These studies involved localized

judgements of depth at, or between, points defined as the vertices of a dense triangular tessellation of

the object surface. Measures included judgements of depth order between pairs of points, magnitude

estimates of local slant/tilt of surface patches using a "gauge" figure probe (see Figure 21), and

magnitude estimates of surface cross-section (depth relief of predefined subset of points on the

surface). Perceived global surface shape for the object was geometrically inferred by interpolating these

local depth judgements to obtain a depth map (e.g., Figure 21). Judged depth relief decreased roughly in

the following order: (1) Binocular viewing of a real or stereoscopic images of the object (2) monocular

viewing of a picture or monocular viewing of a real object (3) Binocular viewing of a picture   This is also

the order of informal reports of reduction in the impression of stereopsis in these studies (as well as in

our own results).

-Figure 21-

However, interpretation of this data is complicated by the fact the global perceived surface shape was

not directly measured, but derived from localized judgments.   One possibility is that these tasks either

partially or fully tap into representations different from those driving the perception of global

surface/object shape. Different depth judgement tasks are thought to access different aspects of depth

representation or levels of processing (Koenderink and van Doorn, 2003; Glennerster, Rogers &

Bradshaw, 1996). For example Di Luca, Domini & Caudek (2010) found that judgements of curvature,

local slant and depth separation between points on the same surface were not metrically consistent. In

the Koenderink and collaborators' studies, it is possible that their local magnitude estimation judgments

require access to absolute scaled depth relations in a way that direct global judgements of 3-D shape or

orientation do not.   If this is true, the more ambiguous absolute depth values becomes (fewer cues,

and/or less reliable distance information) the greater the regression of local depth estimates to the

frontal plane, and the flatter the global depth relief implied by those local judgments. This can also

explain why, despite differences in derived depth relief across viewing conditions and individual

observers, spatial integration of local judgments reveal an internally consistent *global* 3-D shape, and

important correlations and internal consistencies still obtain between viewing conditions and subjects

(Koenderink et al., 1994, 1995).   It is also plausible that some types of depth judgements may partially

access both types of representations. The current proposal which draws a distinction between

representations of relative and absolute (egocentrically scaled) depth therefore suggests a way to

reconcile the results of Koenderink and collaborators with more recent evidence for, and our informal

experience with, global judgements of 3-D shape in pictures.

8.3 The dissociation between stereopsis and the magnitude of perceived absolute depth

While pictures reveal an important dissociation between perceived relative depth magnitude and the

impression of stereopsis, viewing real scenes reveals a dissociation between the perception of absolute

depth values and stereopsis.   The impression of stereopsis appears to diminish for farther viewing

distances even when absolute depth magnitudes are significantly larger (Figure 14; Hildebrand, 1907;

Ogle, 1950; Tscherning, 1905).   Since the cue coherence/depth magnitude account links the degree of

stereopsis directly to depth magnitude, it cannot account for this dissociation.   The disparity hypothesis

can only explain this if it links stereopsis directly to the size of disparities and not scaled depth.   An

alternative view of depth-cue integration suggests that perceived depth magnitude itself is directly

related to the signal-to-noise ratios (reciprocal of CVs) of the individual image signals (Domini & Caudek,

2009; Domini, Caudek & Tassinari, 2006). Specifically, perceived depth magnitude is determined by the number of Just-Noticeable-Differences (JNDs) constituting that particular depth interval.   The psychologically perceived magnitude of depth under this account is simply the sum of JNDs up to the physical stimulus value, suggesting that depth magnitude is not directly linked to physical stimulus magnitude but the relative precision of individual signal(s) at that magnitude.   While the current hypothesis links precision to a *qualitative* attribute (stereopsis), there does appear to be a 'magnitude' or 'strength' to the impression of separation that stereopsis engenders.   Relating psychologically perceived magnitudes directly to precision (signal-to-noise ratios) rather than to physical stimulus magnitudes *per se* (Domini & Caudek, 2009) , is broadly sympathetic to the proposal put forward here and could plausibly account for multiple ways in which depth 'magnitude' is phenomenally perceived.

8.4 The phenomenology of stereopsis: tangibility, immersion and realness.

Reports of naïve and expert observers alike suggest that stereopsis is associated with several specific qualitative characteristics (Barry, 2009, Ogle, 1950, Vishwanath & Hibbard, 2013; Wheatstone, 1838). While the different qualitative characteristics are no doubt tightly coupled, it is useful to distinguish them on the basis primary, secondary and tertiary perceptual qualities (Metzger, 1941; cited in Albertazzi, 2011 & 2012).

The primary qualitative attribute of stereopsis appears to be the sense of 'real' separation in depth between points on an object/part or between objects/parts themselves; the former yielding the characteristic visual impression of 'solidity', and the latter yielding the impression of a 'real' negative space. Additionally, there are other primary visual qualities not associated with the sense of depth that are also reported in the presence of stereopsis: a sense of clarity and visual sharpness, a more enhanced impression of colour and colour variation, and a more enhanced impression of material qualities such as glossiness, shininess, roughness etc. (Barry, 2009; Vishwanath & Hibbard, 2013).

The two secondary phenomenal characteristics associated with stereopsis are the sense of tangibility and the sense of spatial immersion. Tangibility can be thought of as the impression of the manipulability of a material object, while spatial immersivity can be thought of as the impression of the capacity to move through a palpable negative space. Both these effects are incorporated in a vivid sense of protrusion where a tangible solid object reaches or looms out through the negative space toward the observer (Barry, 2009; Vishwanath & Hibbard, 2013). The combined effect of these primary and secondary qualitative aspects is a tertiary qualitative sense of 'reality' or 'realness' to visual objects or scene.

The primary and secondary characteristics described above, all appear related to a sense of the capacity to interact with objects in a 3-D space. Successful motor interaction has a singular prerequisite: reliable egocentrically defined perceptual estimates of direction and spatial scale (size, distance, and absolute depth). While egocentric directions are directly specified in the 'local signs' of the retina, the capacity for motor interactability in 3 dimensions will depend crucially on the availability, consistency, and reliability of sensory information for egocentric distance, scale and depth. The impression of interactability should ideally covary with an understanding of how precisely and definitively spatial relations can be estimated from an egocentric frame of reference, rather than simply the perceived magnitude of a spatial interval. The impression of interactability cannot be derived from an understanding of the veridicality of a perceived depth separation, since there is no basis for determining veridicality until the percept is tested against the somatosensory consequence of the executed action.

By linking stereopsis to the precision of egocentric scaling of visual space, the current hypothesis provides a more plausible functional explanation of the primary and secondary qualitative characteristics associated with stereopsis. The capacity to consciously experience the reliability of derived estimates of egocentric parameters would allow the perceiver to have a direct impression of interactability; specifically, an impression of the *probability of the success of a motor action*. In contrast, if the

impression of motor interactability were generically linked to a representation of relative or absolute

depth *magnitudes* alone, it could encourage suboptimal or erroneous motor planning and programming;

for example, relying on depth estimates that are too imprecise to guide precision movements, or

programming an action to an illusory object in pictorial space. The current hypothesis is therefore

consistent with Michotte's claim that the experience of reality associated with stereopsis is the

impression of the potential for an object to be manipulated, while the perception of 3-Dimensionality or

relative depth structure is simply related to the capacity to visually match the 3-D shapes of objects

(Michotte, 1948).

9. Summary: Explaining Broader Links between Stereopsis and the Quantitative and Qualitative Perception of Depth.

Important aspects of stereopsis related to the perception of both depth magnitude and depth

phenomenology appear inconsistent with existing hypotheses but can be more readily explained on the

basis of the current hypothesis as summarized in Table 6 below.

Table 6. The Capacity for Five Different Hypotheses to Explain Broader Perceptual Aspects

Associated with Stereopsis.

	Binocular Disparity Hypothesis	Visual Parallax Hypothesis	Cue Conflict/ Depth Magnitude Hypothesis	Picture-Awareness Hypothesis	Absolute Depth Scaling Hypothesis
Apparent dissociation between stereopsis and perceived magnitude of absolute depth separation	○	○		○	•
Apparent dissociation between stereopsis and perceived magnitude of relative depth separation			†	○	•
Perception of metric 3-D shape despite lack of stereopsis in pictures					•
Phenomenal characteristics of stereopsis: tangibility , immersion and realness					•

Note: Filled circled represents conditions that can be explained by each respective hypothesis while outline circle represent conditions that can be only partially explained.

10. Conclusion

A unified understanding of the phenomenon of stereopsis and the validation of the theory

presented here will no doubt require sustained empirical investigation. However, it is clear that current

models are clearly inadequate for attaining a full understanding of the complex nature of the human

experience of the 3rd dimension. The present synthesis adds to the work of several researchers (e.g.,

Cutting, 2003; Koenderink & van Doorn 2003; Todd, Oomes, Koenderink, & Kappers 2001; Domini &

Caudek, 2009; Loomis et al., 2002; Svarverud, Gilson & Glennerster, 2012) that challenges the idea that

the visual system produces a single veridical Euclidean representation of the external world, and that it

is this world that we more or less "see" or use to make judgements. However, the current theory is not

in contradiction to the general principles of conventional depth-cue integration models; namely, that

visual mechanisms must yield efficient visual representations and reflect statistically optimal visual-

motor calibrations given the underlying uncertainty and noise in sensory signals. Where the current

theory departs from conventional models, is their assumption that the goal of the visual system is to

"recover" the veridical metric depth structure of the visual scene. Instead, visual mechanisms, and the

perceptual and phenomenal characteristics they produce, are more likely geared toward yielding

representations that help guide the organism effectively for different survival skills; representation that

are in important ways perceptually and physiologically distinct. A more unified understanding of these

different representations in humans in terms of both quantitative and qualitative aspects will be crucial

for the effective design of 3-D media and virtual reality (Loomis & Knapp, 2003)

Finally, the theory implies that the very aspect of visual perception that makes things appear 'real' is based on the value of an internal *subjective* parameter. The reliability of egocentric perceptual depth estimates (as described here) does not refer to a property of the external world, but is an internal parameter that specifies the efficacy of the perceptual presentation of visual surfaces for motor interaction. If the theory holds true, it implies that the perceptual quality (stereopsis) that we hold so dearly as giving us a grasp on reality is an entirely phenomenal construct—with no external referent-- and presents some interesting philosophical implications regarding the nature of perception.

11. References

Adams, W. J. & Mamassian, P. (2004). Bayesian combination of ambiguous shape cues. *Journal of Vision*, *4*, 921-929.

Albertazzi, L. (2010). The roots of metaphorical information. In L. Albertazzi, G. van Tonder, and D. Vishwanath eds., *Perception Beyond Inference*. *The Information Content of Perceptual Processes*, 345-390. Cambridge Mass.: MIT Press.

Albertazzi, L. (2013) Experimental Phenomenology. An Introduction. In L. Albertazzi ed. *The Wiley-Blackwell Handbook of Experimental Phenomenology*. *Visual Perception of Shape, Space and Appearance*, 1-36. Chichester: Wiley-Blackwell.

Allison, R.S., Gillam, B.J. & Veccelio, E. (2008) Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, *9*, 1–14

Ames, A. (1925). The Illusion of depth in pictures. *Journal of the Optical Society of America, 10*, 137-148

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

Barry, S. (2011). Teaching cross-eyed people to see in 3D, in Eyes on the brain. *Psychology Today*,

   October 15.

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of*

   Physiology, 211, 599-622

Botvinick, M. & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see*. Nature 391*, 756.

Buckley D. & Frisby J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of

   three-dimensional ridges. *Vision Research, 33*, 919–933.

Bülthoff, H. H. & Mallot, H. A. (1988) Integration of depth modules: stereo and shading. *Journal of the*

   Optical Society of America A, 5, 1749–1758.

Burge, J., Peterson, M.A., & Palmer, S.E. (2005). Ordinal configural cues combine with metric disparity in

   depth perception. *Journal of Vision*, *5*, 534-542.

Claparede, E. (1904). Stereoscopie monoculaire paradoxale. *Annales d'Oculistique 132*, 465-466.

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

Academic Publishers.

Cooper, E.A. & Banks, M.S. (2012) Perception of Depth in Pictures when Viewing from the Wrong

Distance. Journal of Vision, 12, 896

Cutting, J. E. (2003). Reconceiving Perceptual Space. In *Looking Into Pictures: An Interdisciplinary*

Approach to Pictorial Space, edited by H. Hecht, R. Schwartz, and M. Atherton. Cambridge, MA: MIT

Press. 215–238.

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

Di Luca, M., Domini, F. & Caudek, C. (2010). Inconsistency of perceived 3D shape. *Vision Research, 50*,

1519–1531

Domini, F., & Caudek, C. (2009). The intrinsic constraint model and Fechnerian sensory scaling. *Journal of*

Vision, *9*, 1-15

Domini, F. Caudek C., & Tassinari (2006). Stereo and motion information are not independently

processed by the visual system. *Vision Research, 46*, 1707-1723.

Domini, F., Shah R., & Caudek C. (2011). Do we perceive a flattened world on the monitor screen? *Acta*

   Psychologica, *138*, 359-366

Doorschot, P. C. A., Kappers, A. M. L., & Koenderink, J. J. (2001). The Combined Influence of Binocular

   Disparity and Shading on Pictorial Shape. *Perception & Psychophysics, 63* (6): 1038–1047.

Eaton, E. M. (1919). The Visual perception of Solid Form. *British Journal Ophthalmology*, *3*, 349–363.

Eby, D.W & Braunstein, M.L. (1995). The perceptual flattening of three-dimensional scenes enclosed by

   a frame. *Perception*, *24*, 981-993

Ehrsson, H.H. (2007). The Experimental Induction of Out-of-Body Experiences. *Science, 24*, 1048

Epstein, W. (1966). Perceived depth as a function of relative height under three background conditions.

   J. Experimental Psychology, 72, 335-338

Epstein, W., Park, J., & Casey, A. (1961). The current status of the size–distance hypothesis. *Psychological*

   Bulletin, *58*, 491–514.

Erkelens, C. (2012). Perceived slant of rectangular grids viewed on slanted screens. *Perception, 41,* (ECVP
Abstract Supplement), 16

Fisher, S. K., & Ciuffreda, K. J. (1988). Accommodation and apparent distance. *Perception, 17*, 609–621.

Foley , J. M. (1977). Effect of distance information and range on two indices of visually perceived

distance. *Perception, 6*, 449-460.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review, 87*, 411–434.

Frisby, J. P., Buckley, D. & Horsman, J. M. (1995). Integration of stereo, texture, and outline cues during

pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception, 24*, 181 – 198

Glennerster, A., Rogers, B. J. & Bradshaw, M. F. (1996). Stereoscopic Depth Constancy Depends on the

Subject's Task. *Vision Research 36*, 3441–3456.

Gogel, W. C. (1961)

Gogel, W. C. (1963). The visual perception of size and distance. *Visual Research, 3*, 101–120.

Gogel, W. C. (1969a).   Equidistance effects in visual fields*.   American Journal of Psychology*, *82*, 342-349.

Gogel, W. C. (1969b). The effect of object familiarity on the perception of size and distance. *Quarterly

Journal of Experimental Psychology, 21,* 239–247.

Gogel, W. C.   (1971).   The validity of the size-distance invariance hypothesis with cue reduction.

Perception & Psychophysics, *1*, 92-94.

Gogel, W.C. (1990). A theory of phenomenal geometry and its applications. *Perception & Psychophysics*,

48, 105-123.

Gogel, W. & Da Silva J. A. (1987). Familiar size and the theory of off-sized perceptions. *Perception &*

Psychophysics, 41, 318–328.

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

& Psychophysics, 13, 184-292.

Gogel, W.C. & Teitz, J. D. (1977). Eye fixation and attention as modifiers of perceived distance.

Perceptual & Motor Skills, 45, 343-362.

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.

Helmholtz, H. von (1866/1924) *Treatise on Physiological Optics: Vol III.* (Translated by J.P.C. Southall)

New York: Dover

Hibbard, P. B. (2008). Can Appearance Be So Deceptive? Representationalism and Binocular Vision.

Spatial Vision 21, 549–559.

Hildebrand, A. von, (1907). *The problem of form in painting and sculpture.* New York : G. E. Stechert

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*, 1627-1630.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant From Texture and Disparity Cues:

   Optimal Cue Combination. *Journal of Vision 4*, 967–992.

Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant.

   American Journal of Psychology, 54, 21–37.

Howard I. P. & Rogers B. J. (2002). *Seeing in depth: Vol. 2. Depth perception.* Thornhill, Ontario:

   University of Toronto Press.

Kaufman, L. and Rock, I. (1962). The moon illusion. I. *Science*, 136, 953-61

Kaufman, L., Kaufman, J. H., Noble, R. Edlund, E., Bai, S. & King, T. (2006) Perceptual distance and the

   constancy of size and stereoscopic depth. *Spatial Vision, 19,* 439 – 457.

Kenyon, R.V., Ciuffreda, K.J. & Stark, L. (1978) Binocular eye movements during accommodative

   vergence. *Vision Research, 18*, 545–555

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

   1071–1086.

Koenderink, J. J. & van Doorn, A. J. (2003). Pictorial Space. In *Looking Into Pictures: An Interdisciplinary*

   Approach to Pictorial Space, edited by H. Hecht, R. Schwartz, and M. Atherton. Cambridge, MA: MIT

   Press. 239–299.

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., van Doorn A. J. & Kappers A. M. L. (1995). Depth relief. *Perception,* 24, 115 – 126

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2001). Ambiguity and the 'Mental Eye'

   in Pictorial Relief. *Perception,* 30, 431–448.

Koenderink, J., van Doorn, A. J. & Wagemans, J. (2011). Depth. *i-Perception, 2*, 541–564

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

   Press.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and Modeling of Depth

   Cue Combination: In Defense of Weak Fusion. *Vision Research, 35*, 389–412.

Lappin, J. S. & Craft, W. D. (2000). Foundations of spatial vision: From retinal images to perceived
   shapes. *Psychological Review*, 107, 6-38.

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

environment. In L. J. Hettinger & M. W. Haas (Eds.), *Virtual and adaptive environments: Applications,*

implications and human performance. New Jersey, USA: Lawrence Erlbaum. p21–46.

Loomis, J. M. & Philbeck, J. W. (1999). Is the anisotropy of perceived 3-D shape invariant across scale?

Perception & Psychophysics, 61, 397-402.

Loomis, J. M., Philbeck, J. W., & Zahorik, P. (2002). Dissociation of location and shape in visual space.

Journal of Experimental Psychology: Human Perception and Performance, 28, 1202-1212.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of*

Visual Information. New York: W. H. Freeman.

Marshall, J., Burbeck, C., Ariely,D., Rolland, J. & Martin, K. (1996). Occlusion edge blur: A cue to relative

visual depth. *Journal of the Optical Society of America* A, 13, 681–688

Mather G (1996) Image blur as a pictorial depth cue. *Proceedings of the Royal Society, B*, 263, 169-171.

Mausfeld, R. (2003). Conjoint Representations. In *Looking Into Pictures: An Interdisciplinary Approach to*

Pictorial Space, edited by H. Hecht, R. Schwartz, and M. Atherton. Cambridge, MA: MIT Press.

McCready, D. (1965). Size–distance perception and accommodation–convergence micropsia: A critique.

Vision Research, 5, 189–206.

McIntosh R.D. & Lashley, G. (2008). Matching boxes: familiar size influences action programming.

Neuropsychologia, 46(9):2441-4.

Metzger, W. (1941/1963). *Psychologie: die Entwicklung ihrer Grundannahmen seit der Einftihrung des*

Experiments. Dresden: Steinkopf

Michotte, A. (1948). The Psychological Enigma of Perspective in Outline Pictures. Translated in G. Thines,

A. Costall & G. Butterworth (Eds.) (1991) *Michotte's Experimental Phenomenology of Perception.*

Erlbaum: Hillstereopsisle, New Jersey.

Milner, A.D. & Goodale, M.A. (2008). Two visual systems re-viewed. *Neuropsychologia 46*, 774-785.

Mon-Williams, M. & Tresilian, J. R. (1999a). Some recent studies on the extraretinal contribution to

distance perception. *Perception*, *26*, 167-181.

Mon-Williams, M & Tresilian, J. R. (1999b) The Size-Distance Paradox is a Cognitive Phenomenon,

Experimental Brain Research, 126, 578-582.

Musatti, C. L. (1924). Sui fenomeni stereocinetici *Archivio Italiano di Psicologia*, *3*, 105-120.

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

O'Leary, A., & Wallach, H. (1980). Familiar size and linear perspective as distance cues in stereoscopic

    depth constancy. *Perception & Psychophysics, 27*, 131–135.

Ono, M.E., Rivest, J. & Ono, H. (1986). Human Depth Perception as a Function of Motion Parallax *and*

    Absolute-Distance Information. *Perception and Performance*, *12*, 331-337.

Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon.

    Nature, 414, 197–200.

Ooi, T. L. & He, Z. J. (2007). A distance judgment function based on space perception mechanisms –

    revisiting Gilinsky's equation. *Psychological Review*, *144*, 441-454.

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.

Peacocke, C. (2002). Sensation and the content of experience, sense and content: experience, thought

    and their relations, in: *Philosophy of Mind, Classical and Contemporary Readings*, Chalmers, D. J. (Ed.),

    pp. 435–446. Oxford: Oxford University Press.

Philbeck, J. W. & Loomis, J. M. (1997) Comparison of two indicators of visually perceived egocentric

    distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human*

    Perception and Performance, *23*, 72-85

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

Ponce, C. R. & Born, R. T. (2008) Stereopsis**.** *Current Biology*. 18, R845-50.

Predebon, J. (1993). The Familiar-size Cue to Distance and Stereoscopic Depth Perception. *Perception, 22*, 985–995.

Proffitt, D., Rock, I., Hecht, H. & Schubert,J. (1992). Stereokinetic Effect and Its Relation to the Kinetic Depth Effect. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 3-21

Reinhardt-Rutland A. H. (1999). The framing effect with rectangular and trapezoidal surfaces: Actual and pictorial surface slant, frame orientation, and viewing condition. *Perception 28*, 1361–1371.

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspectives and binocular stereopsis. *Nature, 361,* 253–255.

Rogers, B. J., and Graham, M. E. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*, 125-34

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

Saunders, J.A. & Backus, B.T. (2006). The accuracy and reliability of perceived depth from linear

   perspective as a function of image size. *Journal of Vision, 6*, 933-954

Schlosberg, H. (1941). Stereoscopic Depth from Single Pictures. *The American Journal of Psychology, 54*,

   601–605.

Sedgwick, H. A. (1986). Space Perception. In K. Boff, L. Kaufman, & J. Thomas (Eds*.), Handbook of

   perception and human performance* (pp. 1-57). New York: Wiley.

Smith, P. C. & Smith, O. W. (1961).   Ball throwing responses to photographically portrayed targets.

   Journal of Experimental Psychology, 62, 223-233.

Svarverud E, Gilson S, Glennerster A (2012) A Demonstration of 'Broken' Visual Space. *PLoS ONE* 7,

   e33782.

Todd, J. T., Oomes, A. H. J., Koenderink, J. J. & Kappers, A. M. L. (2001). On the affine structure of

   perceptual space. *Psychological Science*, *12*, 191-196.

Tscherning, M.H.E. (1904) *Physiologic optics: dioptrics of the eye, functions of the retina, ocular

   movements and binocular vision*. Philadelphia: Keystone Press

Tye, M. (2002). Visual Qualia and Visual Content Revisited. In David J. Chalmers (ed.), *Philosophy of

   Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.

Ungerleider, L.G. & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, and

    R.J.W. Mansfield (Eds.). *Analysis of Visual Behavior*. Cambridge: MIT Press, pp.549-586.

Viguier, A., Clement, G., & Trotter, Y. (2001). Distance perception within near visual space. *Perception,*

    30, 115–124.

Vishwanath, D. (2005). The Epistemological Status of Vision and Its Implications for Design. *Axiomathes,*

    15, 399–486.

Vishwanath, D. (2007). Is focal blur an absolute cue to depth. *Journal of Vision, 7(abstract),* 845

Vishwanath, D. (2011). Information in surface and depth perception: reconciling pictures and reality. In

L. Albertazzi, G. van Tonder & D. Vishwanath (Eds). *Perception Beyond Inference, The Information*

Content of Visual Processes. Cambridge: MIT Press.

Vishwanath, D., & Blaser, E. (2010). Retinal blur and the perception of egocentric distance. *Journal of*

    Vision, 10, 1–16.

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. & Hibbard, P. (2010). Quality in depth perception: the plastic effect. *Journal of Vision* 10
    (abstract), 42

Vishwanath, D. & Hibbard, P. (2013). Seeing in 3D with just one eye: Stereopsis in the absence of

binocular disparities *Psychological Science*

Vuong, Q. C., Domini, F. & Caudek, C. (2006). Disparity and shading cues cooperate for surface

interpolation. *Perception,* 35, 145 – 155

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

Leonardo, 34, 231–235.

Wallach, H. & O'Connell D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology,* 45

(4): 205–217.

Wallach, H. & O'Leary, A. (1982). Slope of regard as a distance cue. *Perception & Psychophysics*, 31, 145-

148.

Wang, B. & Ciuffreda K. J. (2004). Depth-of-focus of the human eye in the near retinal periphery. *Vision

research*, *44*, 1115-25.

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

Vision, 5*, 834–862

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

Wu, B., Ooi, T. L. & He, Z. J. (2004). Perceiving distance accurately by a directional process of integrating

ground information. *Nature*, 428, 73-77.

Young, M. J., Landy, M. S. & Maloney, L. T. (1993). A perturbation analysis of depth perception from

combinations of texture and motion cues. *Vision Research*, *33*, 2685-2696.

Zimmerman, D. L., Legge, G. E., and Cavanagh, P. (1995). Pictorial Depth Cues: A New Slant. *Journal of

the Optical Society of America A, 12*, 17–26.

11. Figure Captions

Figure 1. Photographs simulating the binocular viewing of a real 3D scene. The central image is the left eye image and the flanking images are the right eye image. Cross or parallel fuse (respectively) the left or right pair of images.

Figure 2. The induction of stereopsis in static single pictures using a monocular aperture. View a photograph of a real scene with one eye through an oval aperture (1 – 1.5 cm wide) located near the eye at a distance of 2-3cm such that it obscures the rectangular boundaries of the image (see Supplementary Figure 1 for a sample image and aperture). The effect is best appreciated if the image is viewed from about 50 cm on a computer monitor and is zoomed-in such that it is about 25cm in width.

Figure 3. (a) simple 3D object consisting of two rectangular planar surfaces connected at a dihedral angle of θ. (b) An observer views the object (shown in profile). The orientation and the absolute depth dimensions of the object with respect to the observer are given by $\emptyset$, d1 and d2. (c) A projected retinal image. Note that the left and right eye images will be slightly different (see Figure 1).

Figure 4. (a) Depth perception in the absence of egocentric distance information. Based on the information in the retinal image, the observer perceives the 3D shape of the object, but does not know how far away or how big the object is. The observer cannot therefore estimate the absolute values d1 and d2. (b) Depth perception in the presence of reliable egocentric distance information (dotted arrow). The observer perceives the distance and size of the object. The distance information can be used to scale the depth cues in the retinal image to derive the absolute depth values d1 and d2.

Figure 5. Stereopsis and the precision of scaled depth estimates according to the Absolute Depth Scaling Hypothesis. The left panel is the retinal image. In the top right panel, distance is not known and estimates of absolute depth and size are entirely ambiguous; therefore the observer does not perceive any impression of stereopsis. In the bottom right panel, a reliable estimate of distance is available, yielding a precise estimate of absolute depth and size; the observer therefore perceives a strong impression of stereopsis. In the middle right panel, the observer has a less reliable estimate of distance and a less precise estimate of absolute depth; and therefore perceives a weaker impression of stereopsis. Note that in all three cases the observer continues to perceive the same relative depth structure (3D shape)

Figure 6. The distinction between perceived absolute depth magnitude and the impression of stereopsis. In the upper panel, imprecise distance information specifies a farther object. In the lower panel a very precise estimate of distance specifies a closer object. In the upper panel the observer perceives large absolute depth values, but imprecisely. In the lower panel, the observer perceives much smaller absolute depth values very precisely. According to the Absolute Depth Scaling Hypothesis, the observer will perceive a stronger impression of stereopsis in the second case despite perceiving significantly smaller absolute depth magnitudes.

Figure 7. (From Vishwanath & Hibbard, 2013). Naïve subjects' responses to 7 targets items (white region) and 7 distracter items (grey region) describing visual impression comparing either monocular-aperture viewing (left plot) or binocular viewing of stereoscopic anaglyphs (right plot) to the control condition (binocular viewing of single pictures). The two plots are for two different groups of subjects (n =16 each). Grey squares are median responses and black circles are mean responses. Black error bars are

SEMs, and light grey error bars represent the range of responses. Descriptors refer to the test viewing

condition unless indicated otherwise.

Figure 8. (From Vishwanath & Hibbard, 2013). Ratings of perceived difference in the impression of

stereopsis for different pairs of viewing conditions (indicated at left). Ratings were made in relation to

the reference comparison (binocular versus monocular-aperture viewing) which was given an arbitrary

value of 5 units. Zero indicates no perceived difference between viewing conditions in the pair. Error

bars are SEMs. Light grey regions roughly indicate ratings expected based on predictions of the

binocular cue-conflict explanation for monocular stereopsis. Note that these are not quantitatively

derived, but merely used as a visual aid to indicate expected outcome for each pair.

Figure (From Vishwanath & Hibbard, 2013) Curvature-in-depth settings for binocular and monocular-

aperture viewing of horizontally (left) or vertically (right) oriented textured elliptical hemi-cylinders. The

curvature-in-depth is expressed as the axis ratio (a/b) of the elliptical cross section, where 'a' represents

the depth dimension and 'b' represents the vertical or horizontal dimension (depending on cylinder

orientation). Data are averaged over 7 naïve subjects. Shaded regions indicate +/- 1 SEM.

Figure 9. A picture of some generic objects. The 3D shapes are clearly perceived but the distance and

scale of the objects from the observer are ambiguous. Size and distance may be cognitively inferred by

assuming that the horizon is at the eye level of a standing observer. However, assuming that the picture

was taken from a higher or lower elevation alters the inferred distance and size.

Figure 10. Monocular stereopsis in pictures explained on the basis of the Absolute Depth Scaling

Hypothesis. (a) An observer views a picture of the object shown in Figure 3. The picture surface is visible

due to binocular disparity cues and the visible picture frame. Distance information from binocular

convergence, vertical disparity and accommodation specify the distance of the visible picture surface

(dotted arrow). The observer therefore correctly perceives the size and location of the picture surface.

The observer also perceives the 3D shape (depth ratios) of the object depicted in the picture, since the

retinal image is the same as shown in Figure 3C. However, since there are no cues within the picture to

specify the distance of the object, the observer cannot derive its absolute size or depth values. (b) The

same picture viewed monocularly through an aperture. The picture surface is no longer visible, since

binocular disparity and the visibility of the picture frame is eliminated. In the absence of a visible picture

surface, distance information deriving from residual cues such as accommodation (thin dotted arrow) is

postulated to be assigned to the pictorial object allowing a derivation of size and absolute depth values.

Figure 11. Demonstration of perceived miniaturization of a pictorial scene under monocular-aperture

viewing. View the top image through a monocular aperture as described in Figure 2. Accompanying an

impression of stereopsis, a significant change in perceived distance and scale of the scene is observed.

Naïve subjects have reported that the scene appears miniaturized. The effect is similar to the so-called

tilt-shift miniaturization effect induced by the addition of depth-of-field blur as shown in the bottom

image (see text).

Figure 12. (a) Rank order of the perceived strength of the impression of stereopsis for four different

viewing conditions: Binocular and Monocular viewing of a real object (R-B, R-M), monocular-aperture

and binocular viewing of a picture (P-MA, P-B). Black bars are medians and grey bars are means. Error

bars indicate the range of responses.   (b) Ratings of the strength of the impression of stereopsis for the

same four viewing conditions by a different group of subjects. Thick black horizontal bar indicates the

median and thick grey horizontal bar the mean.

Figure 13. Photograph of real scene to illustrate the diminishment of stereopsis with egocentric distance. Arrows indicate approximate partition of the visual scene in terms of Cutting & Vishton's (1995) definition of egocentric space   Under binocular viewing of such real scenes, stereopsis is most clearly experienced for objects in the foreground (e.g. the lampshade) and objects within so-called action space (e.g. the shed, car and nearby hedges and shrubs). Stereopsis is present, but appears diminished at distances beyond the stone house (>50m). For regions near or beyond the white house (vista space, >500m), stereopsis is not usually discernible (Ogle, 1950; Tscherning, 1904). This change in the impression of stereopsis with distance is best appreciated by comparing depth impression under monocular and binocular viewing of a similar real scene.

Figure 14. Judgments of the strength of the impression of stereopsis in real scenes at different distances operationalized as the magnitude of the difference in depth impression between monocular and binocular viewing. (a) Ratings for a street scene viewed from ground level. Bold values on abscissa indicate viewing distance to nearest aspect of the object(s); values in square brackets indicate the total depth interval of the object(s) in the region of interest (ROI) which subtended about 6-7 degrees visual angle. Thick black horizontal bar indicates the medians and black crosses indicate single outlier values. Means calculated with all values are indicated by thick grey horizontal bars (n=9). (b) Same as left graph for a different group of subjects and scenes. Reference comparison was a flowering shrub viewed at 1m. Subjects viewed the scene from ground level or from a parapet at a height of 8m (light grey region). (c) Examples of scenes and ROIs for data in (b).

Figure 15. (a) Plots of disparity magnitude as a function depth separation from the point of fixation for 10 fixation distances (.3m – 500m; log-log scale). Each curve plots the disparity value of a point located in front of or behind the point of fixation (for each curve, the fixation distance is the point where the

curve touches the abscissa (zero disparity)]. Disparities are indicated in arcsecs, arcmins and arcdegrees

of visual angle. Note the difference in the values of the ordinate in the left and right graphs. (b) Plot of

absolute depth beyond fixation as a function of fixation distance (log-log scale) for 9 different values of

absolute disparity (20 arcsecs – 90 armins; indicated at the top of the graph).

Figure 16. Plots of estimated coefficients-of-variation for absolute depth judgments between two points

as a function of fixation distance (assuming fixation is on the near point) for 9 different disparity values,

each represented by a red curve. Both left and right plots are derived assuming Gaussian distributions

for disparity and egocentric distance estimation. Both graphs were derived assuming standard deviations

for disparity detection based on Blakemore (1970). Left graph uses standard deviations for distance

estimation assuming coefficient-of-variations for distance estimation that increase linearly from 2% to

15% (from 30cm – 200m). Right graph uses standard deviations for distance estimation assuming

constant coefficient-of-variations of 5% for all distances. Note that in both graphs, abscissa is plotted in

log units and ordinate in linear units.

Figure 17. Images of a rectangular slanted object viewed from different distances. The object is 3 x 1.5

units tall and the distance of the object from the observer (optic axis on centre of object) is 10, 20, 40

and 80 units respectively. With increasing viewing distance both the size and convergence angles

become smaller, reducing the reliability of the perspective scaling cues that indicate the quantitative

depth separation of the left and right edge of the object.

Figure 18. Illustration of the reduction in the impression of stereopsis with distance and its relation to

the Absolute Depth Scaling Hypothesis (a) An observer views a near object at a distance of about 1m and

a far object at a distance of about 50m (distances indicated are not to the scale of the depicted human

figure).   (b) Viewing the scene from ground level, a binocular observer is able to estimate the size and

absolute depth values of the near object with a high degree of precision and obtains a strong impression

of stereopsis. The scale of the far object is estimated with considerably lower precision (variability

illustrated by the grey instances of the object). It is therefore perceived with a weaker impression of

stereopsis. Note that since the true distance and scale of the object (black instance) is included within

the 'zone of indeterminacy' the observer should not notice any miniaturization. The observer can

cognitively choose a scale within the zone of indeterminacy that matches familiar size (similar to

pictures). (c) The far object viewed from an elevated parapet in the absence of a visible continuous

ground plane.   Distance cues deriving from vergence, vertical disparity and declination from eye level

are unreliable due to the distance. If the declination from eye level, and default distance tendencies

play a role in determining distance, the observer should perceive the object to be located at an

erroneously closer distance (light-grey instances) and therefore smaller than true size (miniaturized). If

the reliability of this distance estimate is better than that in (b) above, the observer should also perceive

a stronger impression of stereopsis.

Figure 19. Top row: three visual configurations specified by retinal cues (slanted planar surface, concave

dihedral, convex curving surface). Affine structure: perceived 3D shape is only specified up to an affine

transformation. The dark dotted lines indicate possible configurations given disparity or motion parallax

information which specify different shapes/orientations at different distances. Relative depth structure:

the perceived 3D shape (depth relief) is uniquely specified , but its scale is not known. Absolute (scaled)

depth structure: The 3D shape of the object (relative depth structure), its distance, size and absolute

depth values are all perceived. Binocular disparity and motion parallax do not specify a unique 3D shape

(relative depth structure), unless scaled by an estimate of distance (arrow 1 on the right). Pictorial cues

such as perspective and shading can, in principle, specify a unique 3D shape even in the absence of

distance information (arrow 2).

Figure 20. A picture that elicits the perception of a 3D object (adapted from Adelson & Pentland, 1996).

The only conventional depth cue specifying 3D shape is shading. Other available cues (binocular

disparity, motion parallax (assuming small head movements), surface microtexture (pixellization),

successive vergence, defocus blur (if the image is sufficiently large) all specify a flat surface (the picture

itself). While there can be bistability between two 3D interpretations, a flat 2D painted-surface

interpretation is seldom perceived, and requires considerable effort. Moreover there is no apparent

difference in perceived shape between monocular and binocular viewing.

Figure 21 (From Doorschot, Kappers & Koenderink, 2001). The stimuli (example shown at left) were

stereoscopic photographs of a real sculpture of a human torso (shown here with a gauge-figure probe in

black). The three images on the right show response surfaces obtained (respectively, left to right) for

disparity simulating viewing a single picture binocularly, normal disparity simulating binocular viewing of

the real object, double the normal disparity (assuming twice the standard inter-ocular distance).

X ||

-Figure 1-

-Figure 2-

A

B

$-d_2-$

θ

ϕ

d_2

C

-Figure 3-

A

B

d_1

d_2
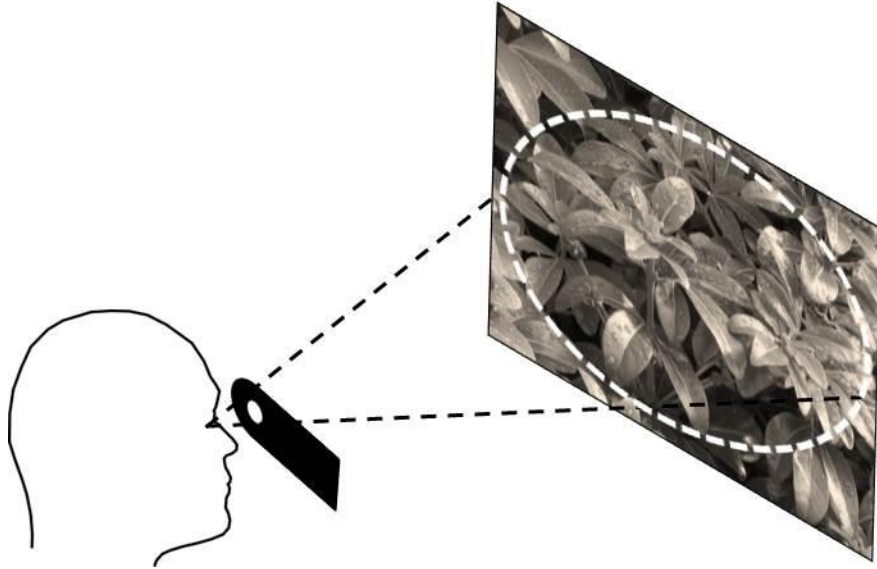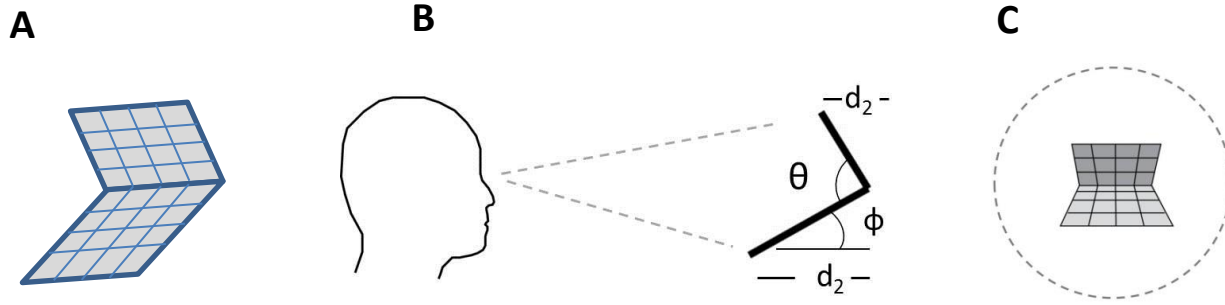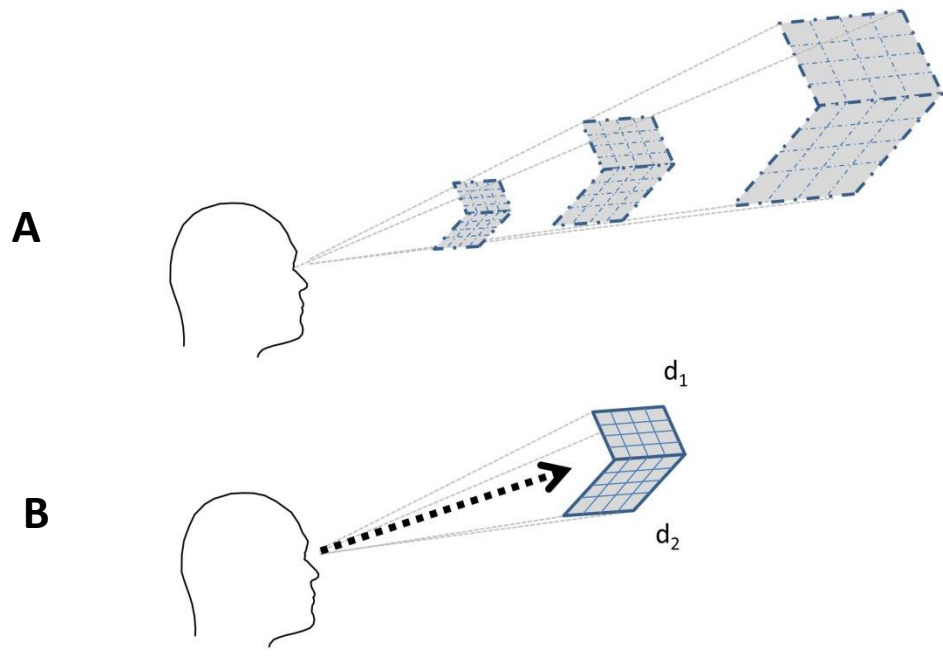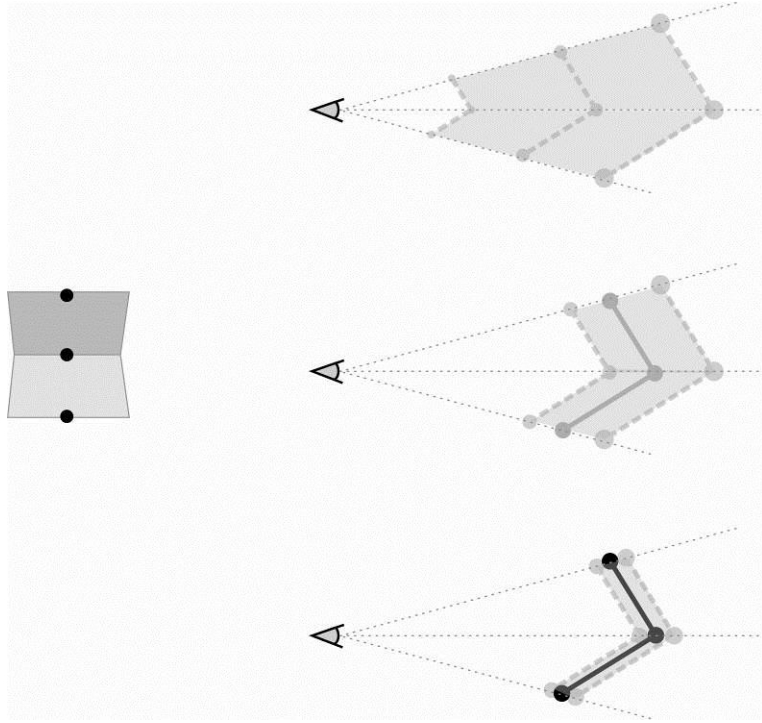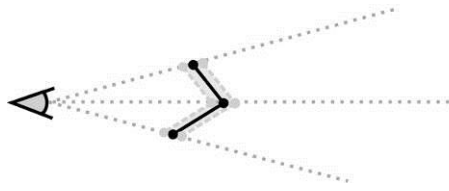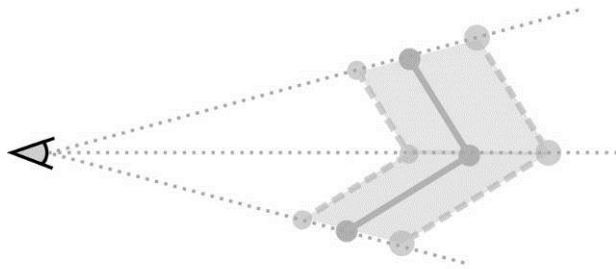
-Figure 4-

-Figure 5-

-Figure 6-

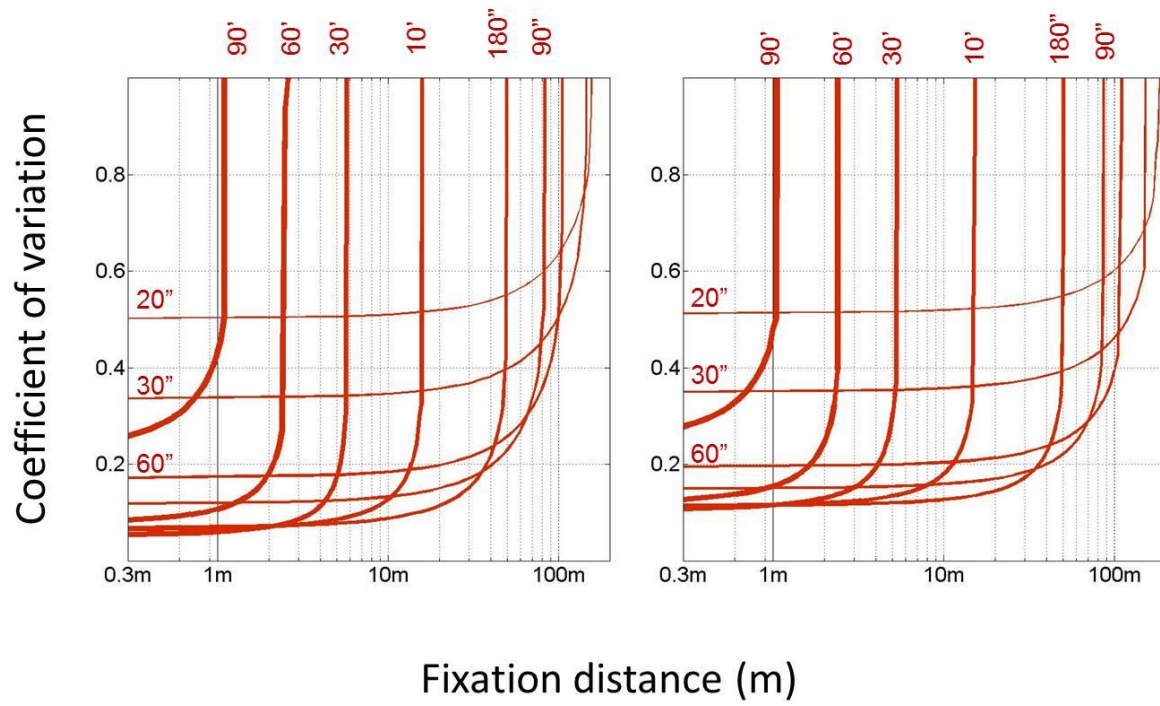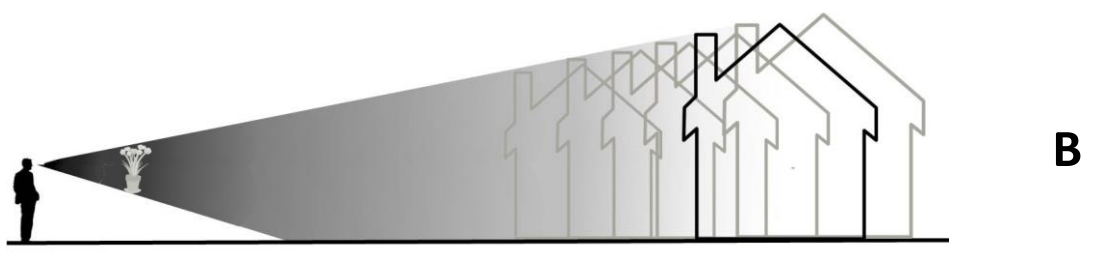Monocular-Aperture Viewing of Single Pictures

Binocular Viewing of Stereoscopic Anaglyphs

Objects stick out of screen
More three-dimensional
Depth is clearer/more defined
More definitive seperation
Objects appear more real
Objects appear touchable
Greater separation

More blurry
More transparent
Objects appear to move
Objects appear distorted
Objects appear distorted in control
Object shapes are different
Objects appear smaller

--- -- - 0 + ++ +++
Disagree   Agree
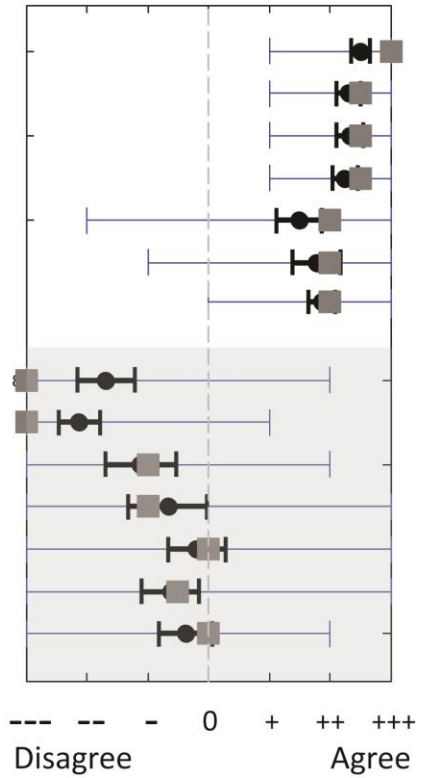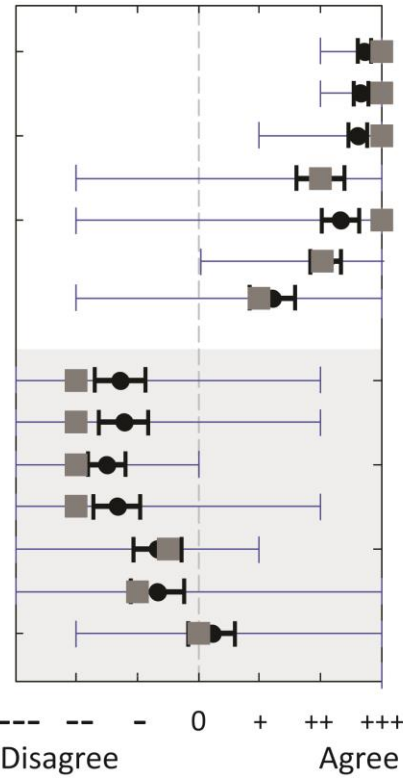
-Figure 7-

-Figure 8-

-Figure 9-

A

B
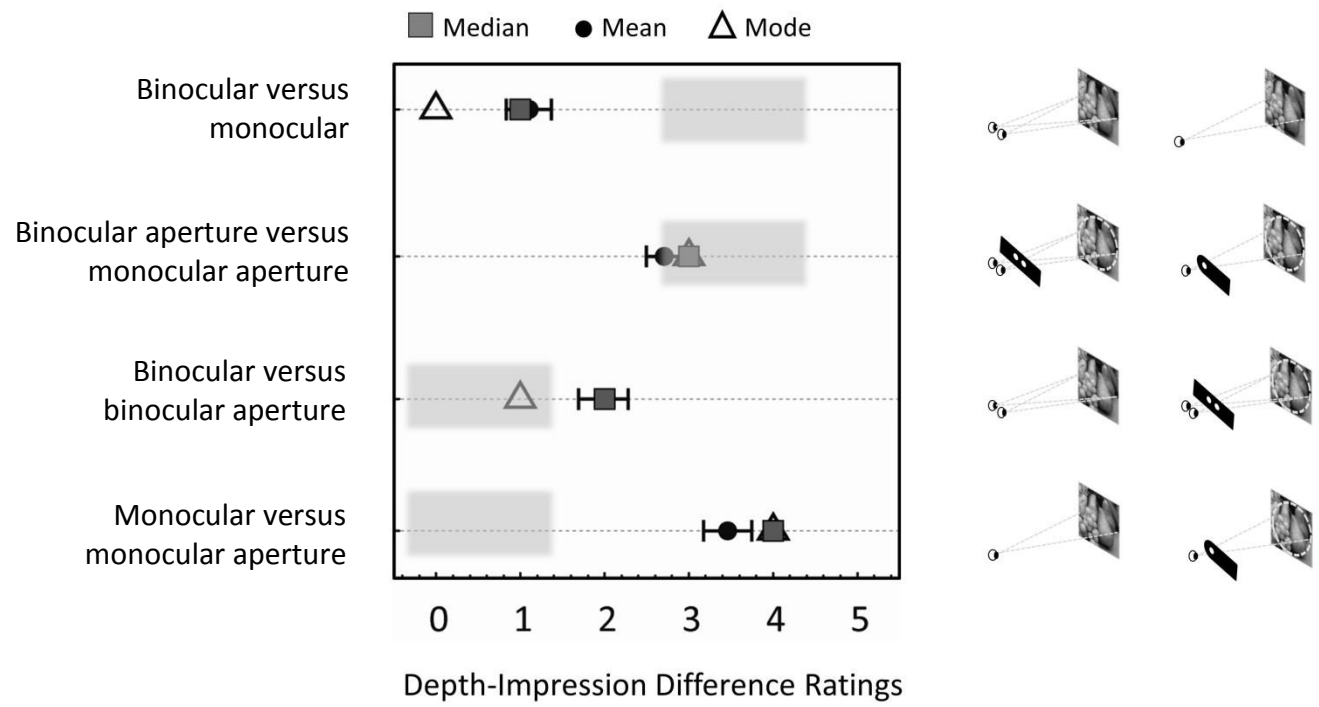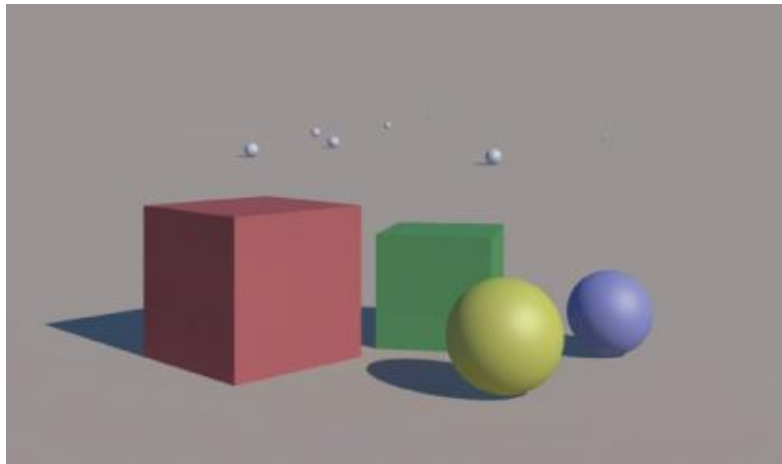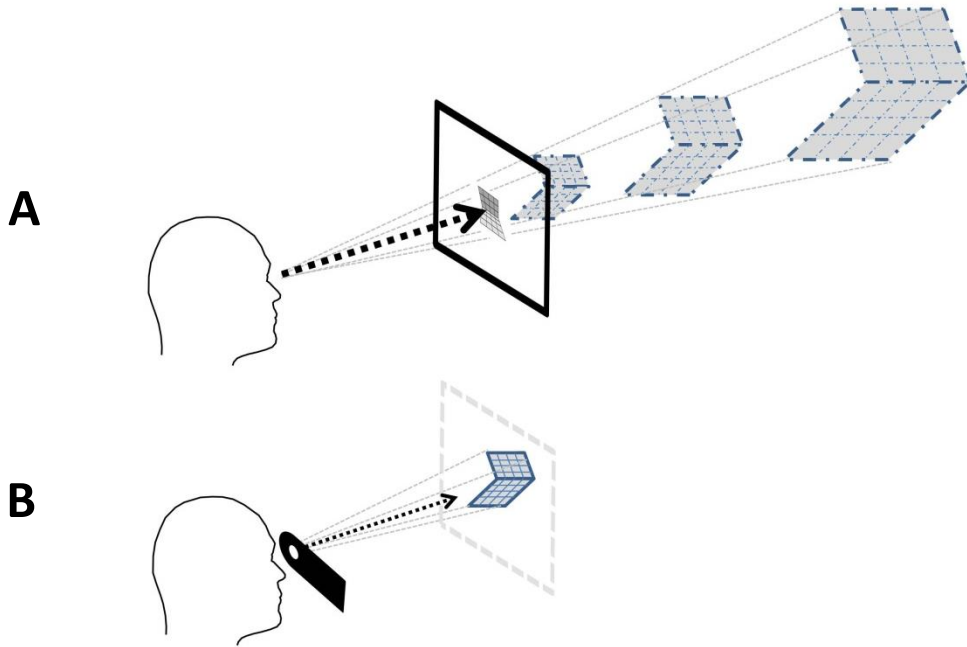
-Figure 10-

-Figure 11-

-Figure 12-

VISTA SPACE

ACTION SPACE

PERSONAL SPACE

-Figure 13-

-Figure 14-

a

Disparity magnitude

2 deg
60'
30'
10'
2'
60"
30"

.3m .5m 1m 2m 5m 10m 50m

10'
2'
60"
30"
10"

10m 20m 40m 100m 500m

Fixation distance (m)

b

90' 60' 30' 10' 180" 90" 60" 30" 20"

Absolute depth (m)

100m
10m
1m
10cm
1cm
1mm

0.3m 1m 10m 100m

Fixation distance (m)
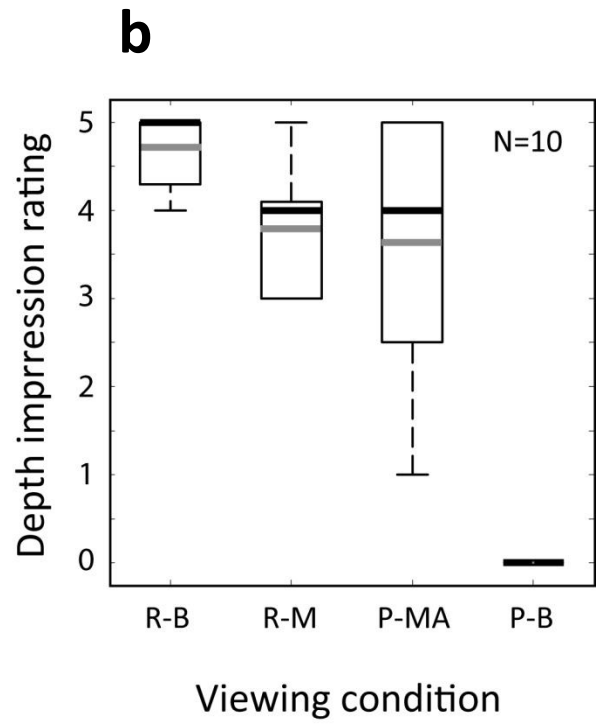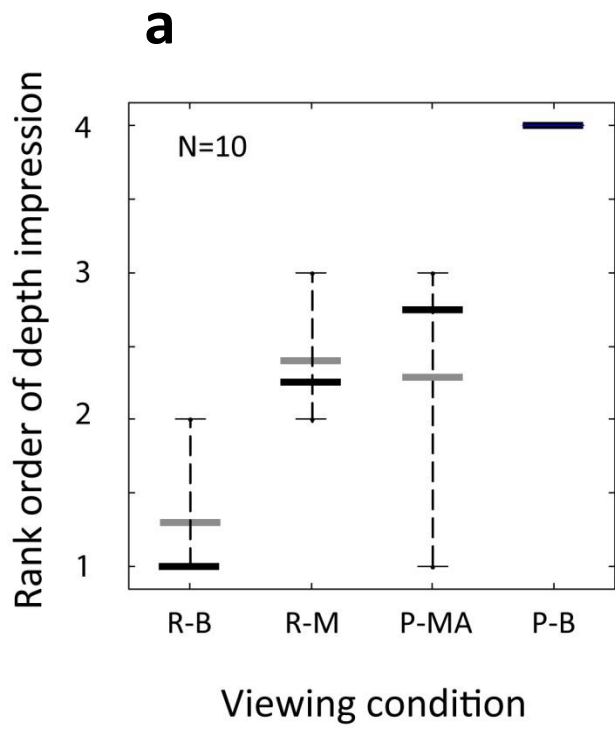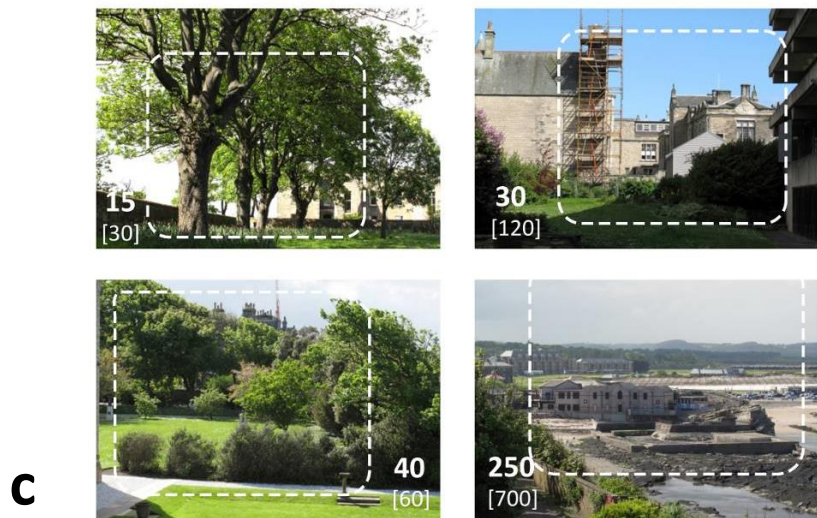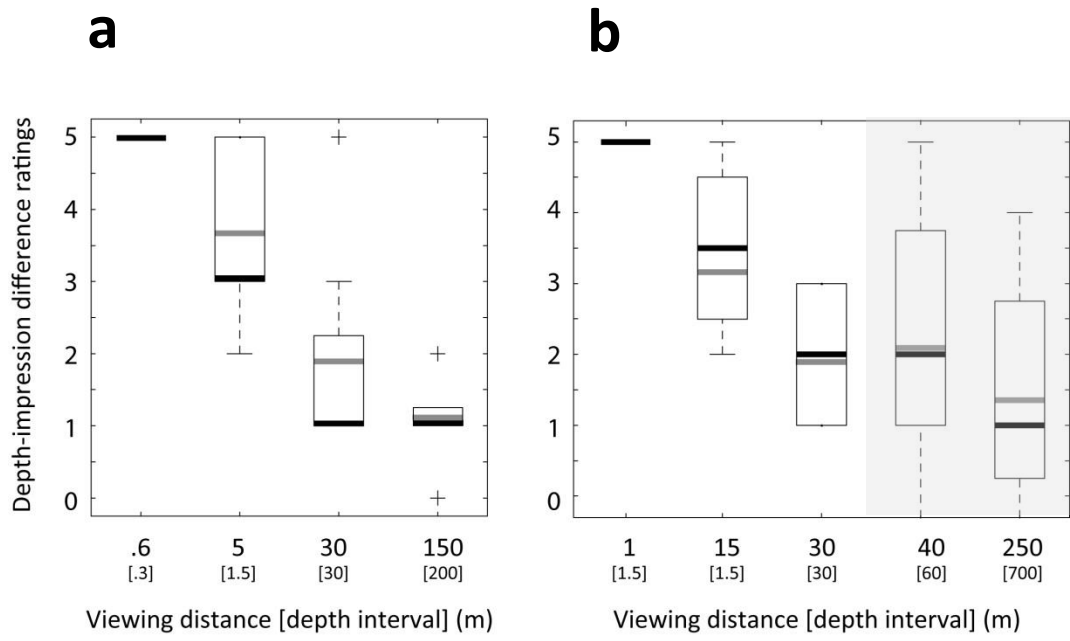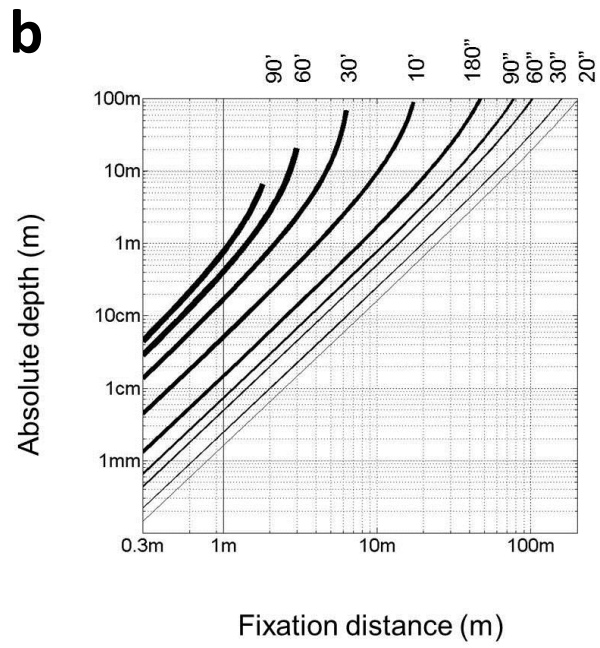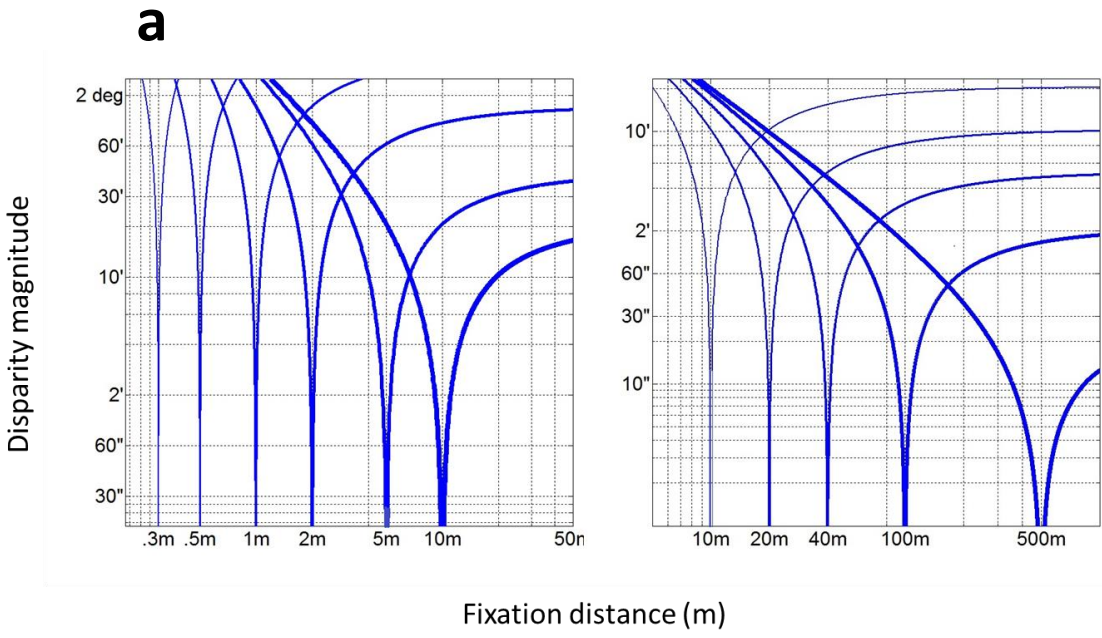
-Figure 15-

-Figure 16-

-Figure 17-

~1m

~50m

A

B

C
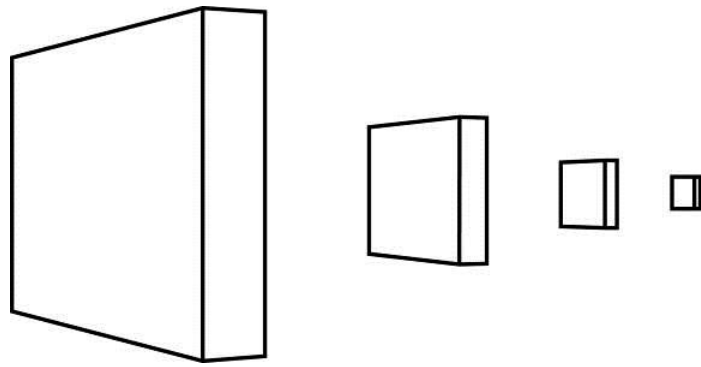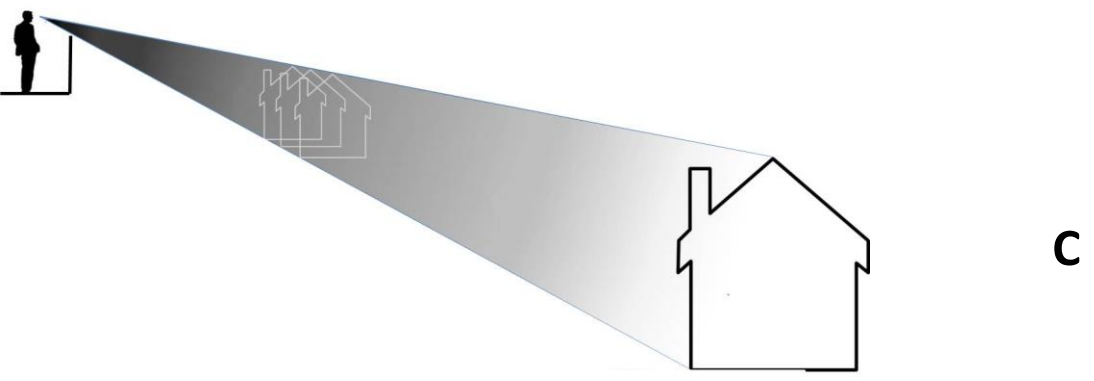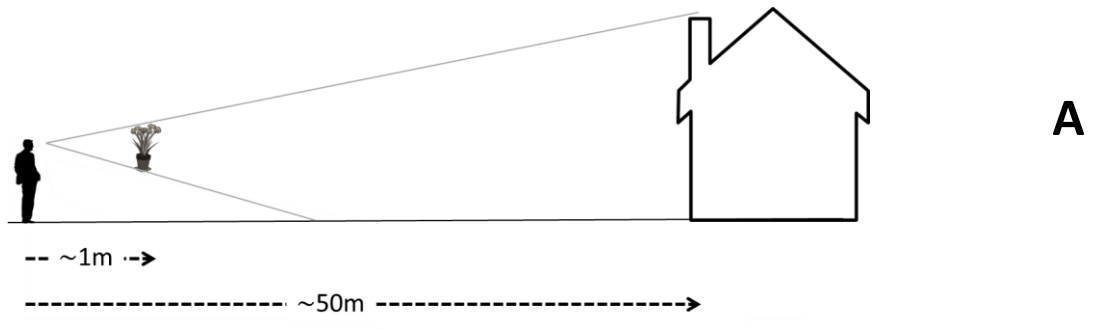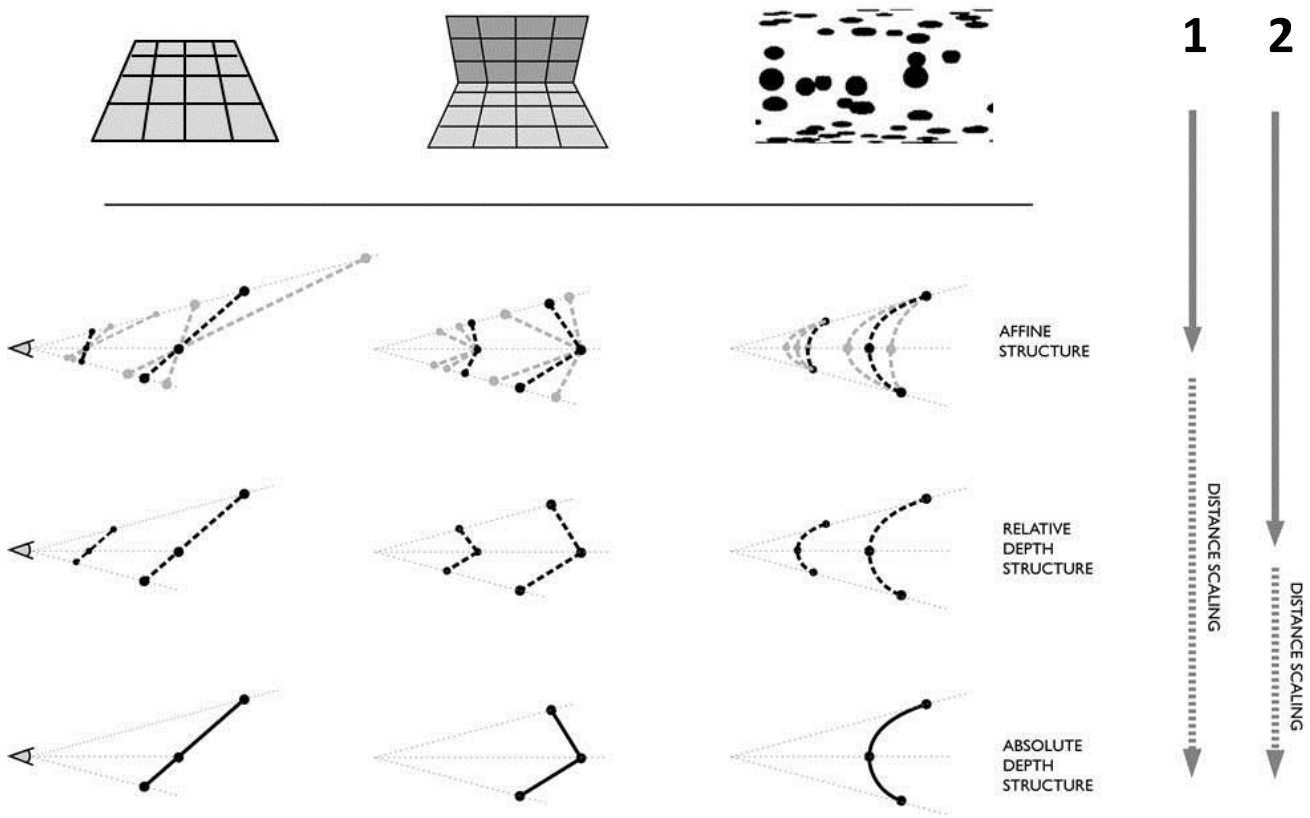
-Figure 18-

-Figure 19-
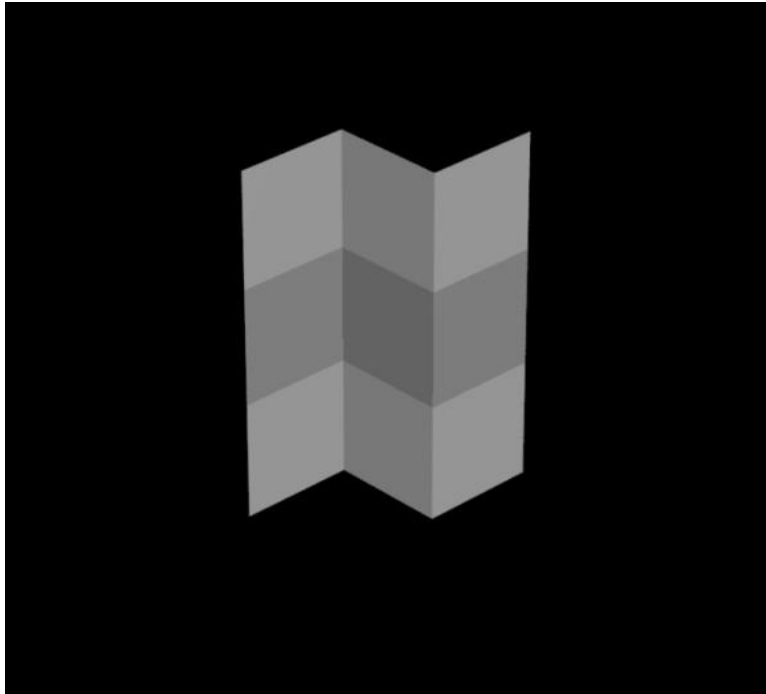
-Figure 20-

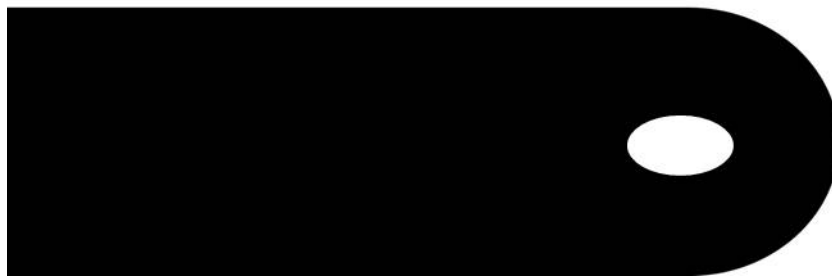*lighting condition*
side lighting
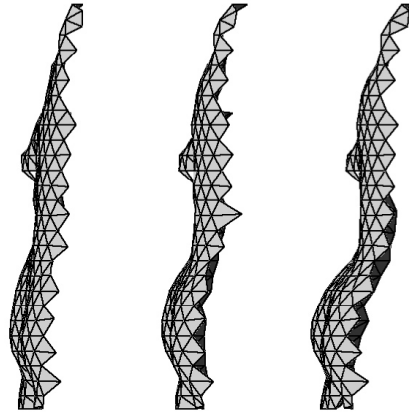
-Figure 21-

a

b

Figure S1. Demonstration of the impression of stereopsis in single pictures using an aperture. (a) Compare viewing the image on a high resolution computer display with both eyes normally and then with one eye through an oval viewing aperture (1.2-1.5 cm). A suitable aperture can be constructed by printing the image in (b) onto card stock and cutting along the edges. The aperture should be placed near the viewing eye (other eye closed) as shown in Fig 2 (main document). The location of the aperture, or the viewing distance, should be adjusted until the frame of the image is completely occluded (see Fig 2) . The image should optimally be about 25cm wide on the display and viewed from 40-50cm. The effect can initially take several seconds to obtain.
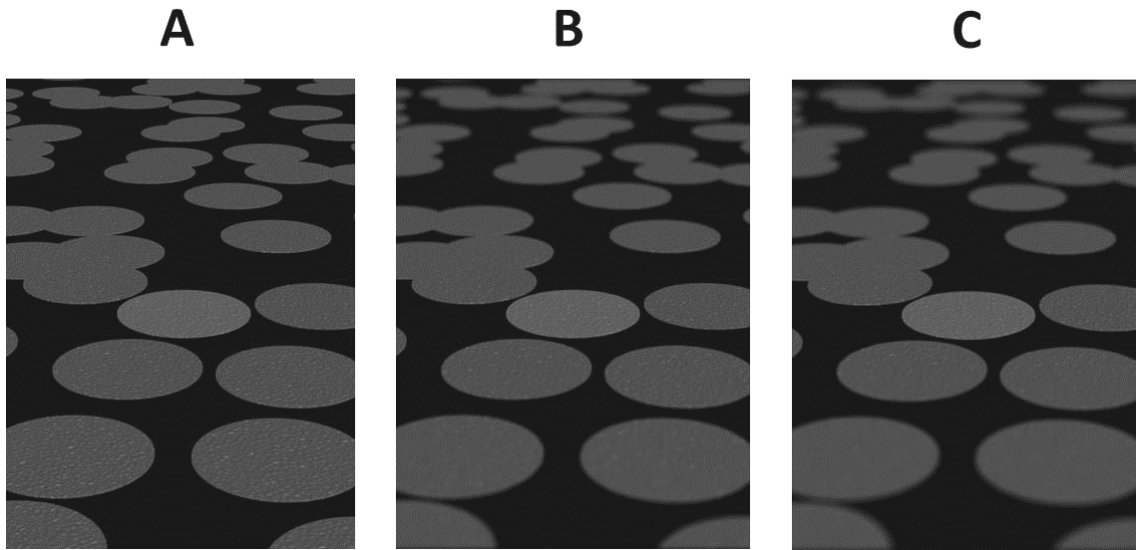
-Figure S2-

Depth-of-focus blur: The pattern of defocus blur for an extended surface or region of space. In A, the textured surface is uniformly sharp. In B and C there is a characteristic gradient of depth-of-focus (DoF) blur increasing from the central texture patch. *Egocentric distance from DoF blur*: The entire surface in *B* appears closer than in *A*; similarly the surface in *C* appears closer than *B* (see Vishwanath, 2010). The images may need to be enlarged on the screen to see these effects.

-Figure S3-