CHAPTER 1

Two Conceptions of Stereopsis

Abstract In this introductory chapter, I outline the two competing conceptions of stereopsis (or depth perception) that have dominated the literature over the last 150 years. The first conceives of stereopsis in purely optical terms, typically as an exercise in inverse optics. By contrast, the second approach argues that optical information from the world is indeterminate until contextual meaning has first been attributed to it. In this book I attempt to advance a purely optical account of stereopsis and I use this introductory chapter to raise the central contention of Chaps. 2 and 3, namely that many of the 'perceptual' phenomena that appear to count against a purely optical account of stereopsis are better understood as post-perceptual cognitive inferences.

Keywords Stereopsis · Visual cognition · Cue integration
Gestalt psychology · Intentionality

If vision is concerned with the perception of objects (Gibson 1950; Strawson 1979), then stereopsis is the visual space in which those visual objects are located, specifically: (a) the volume of space which each object takes up (in

---

The original version of the book was revised: Post-publication corrections have been incorporated. The erratum to the book is available at DOI 10.1007/978-3-319-66293-0_5

P. Linton, *The Perception and Cognition of Visual Space*,
DOI 10.1007/978-3-319-66293-0_1

response to which Wheatstone 1838 coined the term 'stereopsis', the Greek for 'solid sight', also known as the 'plastic' effect), as well as (b) the volume of space between each object (known as the 'coulisse' effect, the French for the space between the flat-panels of stage scenery). According to this definition stereopsis is simply the perceived geometry of the scene or, as Koenderink et al. (2015b) suggest, the perception of three-dimensional space.

This definition of stereopsis contrasts with a number of authors in Philosophy (e.g. Peacocke 1983; Tye 1993) and Psychology (e.g. Hibbard 2008; Vishwanath 2010) for whom stereopsis and the perceived geometry of the scene can come apart. For instance, all four authors argue that whilst closing one eye may reduce stereopsis (i.e. lead to a reduction in our subjective impression of visual depth), it does not affect the perceived geometry of the scene (i.e. the scene itself does not appear to be any flatter).

1 Two Conceptions of Stereopsis

But however stereopsis is defined, the fundamental question is what gives rise to this subjective impression of visual depth? Over the last 150 years there has been an ongoing debate between two schools of thought:

The first school of thought regards our stereoscopic impressions as simply the product of (a) *Optical* cues, such as binocular disparity (the difference between the two retinal images), possibly with the addition of (b) *Physiological* cues, such as accommodation (the focal distance of the eyes) and vergence (the angle between the eyes), but *without* the need for the visual system to (c) attribute *contextual information* or *subjective meaning* to these cues (apart from the limited conceptual apparatus required to construct *any* 3D surface in space). According to this account the perceived geometry of the scene, which we experience as stereopsis, is simply specified by the optical information that we receive from the environment.

This conception of stereopsis is closely associated with Physiological Optics, and has been held at one time or another by Hering (1865; discussed by Turner 1994), Mach (1868, 1886; discussed by Banks 2001), Cajal (1904; cited by Bishop and Pettigrew 1986), Gibson (1950), and Julesz (1960), and still holds sway in contemporary Physiology where 'stereopsis' is often simply defined as:

> The sense of depth that is generated when the brain combines information from the left and right eyes. (Parker 2007)

Layout: A5 HuSSci | Book ID: 455577_1_En | Book ISBN: 978-3-319-66293-0
Chapter No.: 1 | Date: 11 November 2017 14:31 | Page: 3/34

> …the ability of the visual system to interpret the disparity between the two [retinal] images as depth. (Livingstone 2002)

> …the third spatial dimension to be extracted by comparison of the somewhat differing aspects of targets that arise when imaged from two separate vantage points. (Westheimer 2013)

By contrast, I would suggest that this stereoscopic impression of depth is not only present when we view the world with two eyes ('binocular stereopsis'), but also when we view the world with one eye ('monocular stereopsis'). This monocular impression of depth is more commonly attributed to the second conception of stereopsis. The second conception argues that in addition to (a) *Optical* cues (such as binocular disparity), and (b) *Physiological* cues (such as accommodation and vergence), the visual system either (c) also relies upon *Pictorial* cues (such as perspective and shading), whose content is geometrically unspecified until the visual system attributes meaning to them, or (d) treats *all* depth cues (the *Optical* and *Physiological* cues, as well as the *Pictorial* cues) as being unspecified until the visual system attributes meaning to them. And the meaning that the visual system attributes to these depth cues can take one of two forms: (i) *ecologically valid meaning* in the form of *prior knowledge* (typically *natural scene statistics*), or (ii) *subjective meaning* that may or may not correspond to physical reality.

This alternative conception of stereopsis is more closely associated with Cognitive Psychology, and has been held at one time or another by Ibn al-Haytham (c.1028–1038), Berkeley (1709), Helmholtz (1866), Ogle (1950), and Gregory (1966); and is closely related to both the Gestalt Psychology of the early-twentieth Century and Gombrich's (1960) 'beholder's share'. Although Cognitive Psychology has provided us with the leading articulation of this account, and indeed the leading articulation of stereopsis over the last two decades (in the form of Cue Integration, see Landy et al. 1995; Knill and Richards 1996), the argument that stereopsis is specified by the attribution of meaning is broader than Cognitive Psychology (see Albertazzi et al. 2010; Vishwanath 2005). Nonetheless, it is worth emphasising the affinity between this Psychological account of stereopsis and two of the central concerns of 'Cognitive Revolution' of the 1950–1960s:

Layout: A5 HuSSci | Book ID: 455577_1_En | Book ISBN: 978-3-319-66293-0
Chapter No.: 1 | Date: 11 November 2017 14:31 | Page: 4/34

a. Perception as Creative Construction: For Neisser (1967), the 'central problem of cognition' was the fact that visual experience is *creatively* constructed. Indeed, he coined the term 'Cognitive Psychology' to 'do justice … to the continuously creative process by which the world of experience is constructed': 'As used here, the term "cognition" refers to all the processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used. It is concerned with these processes even when they operate in absence of relevant stimulation, as in images and hallucinations'.

b. Centrality of the Mind: This account of perception necessarily presupposes that *the mind* (which Neisser articulated as the software of the brain, in contrast to its physiological hardware) would have a central role in determining the content of perception. Indeed, this had already been a central contention of the 'Cognitive Revolution' in the decade before, in particular the 'New Look' literature that started with Bruner and Goodman (1947). As Miller (2003) explained of the 'Cognitive Revolution': 'We were still reluctant to use such terms as 'mentalism' to describe what was needed, so we talked about cognition instead'.

2 A Recent History of Stereopsis

But whilst Psychology was undergoing a revolution to give the mind a central role in determining perceptual content, stereopsis was about to undergo a transformation in the opposite direction. For instance, Bishop and Pettigrew (1986) draw a distinction between the pre-1960 literature on stereopsis:

> Stereopsis Before 1960: Mentalism Prevails

And the post-1960 literature on stereopsis:

> The Retreat From Mentalism Begins In The 1960s

As Bishop and Pettigrew explain, the pre-1960 literature on stereopsis was an unbroken linage that stretched for a century from Helmholtz (1866) to Ogle (1959). Speaking of the period before 1960, Bishop and Pettigrew observe:

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 5/34

> Before that time it was generally believed that binocular depth perception was based on high-level quasi-cognitive events that took place somewhere in the no-man's land between brain and mind.

For instance, Ogle, the leading authority on binocular stereopsis at the time, wrote in his introduction to *Researches in Binocular Vision* (1950) that depth perception was a synthesis of (a) the retinal stimulation, (b) the physiology and neurological processes by which this retinal stimulation was communicated to the brain, (c) the 'psychic modifications and amplifications' of these 'neurologic "images"' by past visual, auditory, and tactile experiences, and (d) the modifying effects of attention and the motivations of the individual.

Indeed, as late as the 1950s some Gestalt Psychologists still argued that stereopsis was a *purely* Psychological phenomenon (Ogle 1954 cites Tausch 1953, and Ogle 1959 cites Wilde 1950). Although Ogle rejected this extreme position (Ogle 1954), he nonetheless insisted that the *meaning* attributed to pictorial cues by the visual system could modify or even dominate the stereoscopic impression from binocular disparity. For instance, Ogle (1959) distinguishes between (a) stereopsis from binocular disparity, which he regards as automatic, and therefore *meaningless*, and (b) 'empirical clues' such as perspective, which have been made *meaningful* by experience. And he goes on to conclude:

> …it is to be expected that in those surroundings that have been artificially produced to provide a conflict between stereoscopic stimuli and empirical factors, the meaningless stimuli may be suppressed by the meaningful, that is, by the perceptions from the empirical motives for depth.

This passage was heavily influenced by Ogle's mentor at Dartmouth, Adelbert Ames Jr. (1951, 1955). Indeed, Ogle advances Ames' Window (where a trapezoid window frame constructed to look like a rectangular frame seen in oblique perspective appears to change direction as it is rotated) as just such an instance of *meaningful* empirical cues (in this case perspective) dominating *meaningless* binocular disparity.

But the impetus for stereopsis' anti-Cognitive Revolution in the 1960s was not Ogle's (1959) suggestion that binocular disparity could be modified by pictorial cues, but instead his insistence that depth could not be extracted from binocular disparity until figure-ground meaning had first been attributed to both of the retinal images:

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 6/34

> We must stress the importance of contours, those lines of demarcation between the 'figure' and the 'background.' In every case stereoscopic depth depends on the disparity between the images of identifiable contours.

It was in response to this claim that Julesz (1960) created the Random-Dot Stereogram (see also Aschenbrenner 1954). Julesz had been a radar engineer in the Hungarian military where the practice had been to use stereo-images (two images taken from different perspectives) to break camouflage in aerial reconnaissance: the camouflaged object was indiscriminable until the images were fused stereoscopically, at which point the object would jump out in vivid depth. Julesz therefore hypothesised that stereopsis must *precede* contour recognition, and he invented the Random-Dot Stereogram as a form of 'ideal camouflage' to prove this very point: comprised of two images of apparently randomly distributed dots, the contours of a hidden object are encoded in the differences between the images (rather than the individual images themselves), and yet the hidden object emerges in vivid depth when they are fused (Fig. 1).

The Random-Dot Stereogram not only revolutionised the understanding of stereopsis in Psychology, it also had a significant impact upon Neurophysiology by inspiring the search for 'disparity selective neurons' that could track these differences between the two images (see Pettigrew 1965). Indeed, as Cumming and Parker (1999) observe, the discovery of disparity selective neurons by Barlow et al. (1967) and Nikara et al. (1968) would further conflate stereoscopic depth perception and binocular disparity.

But the problem with equating stereopsis and binocular disparity in this way is the implication that monocular vision lacks this subjective impression of depth. For instance, in his Ferrier Lecture on Stereopsis, Westheimer (1994) insisted that 'real stereo sensation is absent with monocular viewing', whilst Parker (2016) appears to suggest that 'a direct sense of depth' only emerges with binocular vision. Although this position continues to have adherents in Physiology, and is attractive to others on purely experiential grounds (see Sacks 2006; Barry 2009; and Sacks 2010), by the mid-1990s it had come to be rejected by Cognitive Psychology. As Landy et al. (1995) observe, in what is arguably the most influential paper of this period, if you close one eye the world does not suddenly become flat. The implication being that stereopsis is not just a

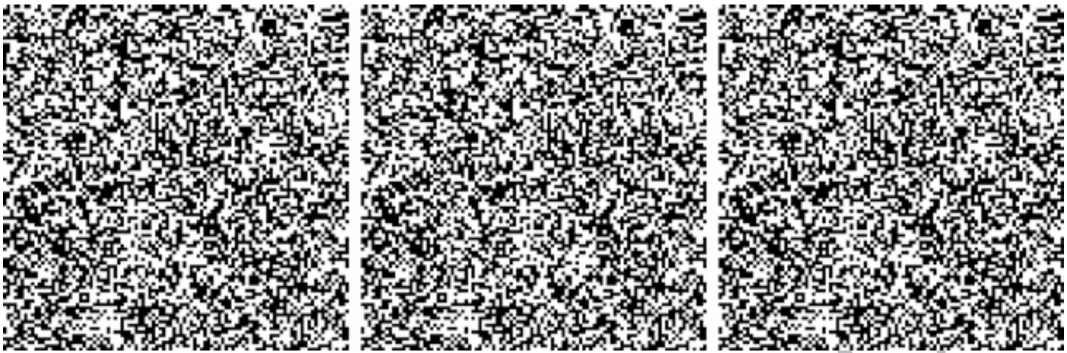

Fig. 1 Random-Dot Stereogram proposed by Julesz (1960). A square appears in stereoscopic depth when you cross-fuse the left and central image (by focusing on a point in front of the image) or parallel-fuse the central and right image (by focusing on a point behind the image) in spite of the fact that there are no contours demarcating the square.

product of binocular disparity but also monocular cues to depth such as shading, perspective, and occlusion. Over the last couple of decades this observation has been explored in two distinct ways:

1. Cue Integration: The first strand in the literature marks a return to Ogle's (1959) observation that monocular depth cues can modify the stereoscopic impression from binocular disparity. As Landy et al. (1995) and Knill and Richards (1996) documented in the mid-1990s, when faced with multiple sources of depth information the visual system appears to combine this information into a single coherent percept either by integrating the information linearly (if the sources of information are only mildly in conflict) or by down-weighting or excluding apparently aberrant sources of information (when the conflicts are large).

As we shall see in Chap. 2, this process of Cue Integration has been systematically tested and confirmed in the literature using cue-conflict stimuli. I do not seek to challenge most of these results. Instead, what I do seek to challenge is the *interpretation* that has been given to these results. To explain why, we should recognise that there is at least one part of the Cognitive Revolution that, some 60 years on, still feels incomplete: as Miller (2003) observed, the Cognitive Revolution was a reaction against the excesses of Behaviourism, according to which *perception* was equated with *discrimination*; *memory* was equated with *learning*; and *intelligence* was equated with *what intelligence tests test*. And yet, even to this day, the tendency to conflate *perception* with *discrimination* still persists. So I argue in Chap. 2 that many cue-conflict experiments appear to reflect their subjects' (mistaken) *judgements* or *evaluations* of their visual experience, rather than the depth that is actually perceived. The same, I argue, also appears to hold true of cue-conflict illusions such as the hollow-face illusion and Reverspectives, and I suggest that they are better thought of as *delusions* (false judgements) rather than *illusions* (false percepts).

2. Monocular Stereopsis from a Static 2D Image: The second implication of depth from pictorial cues is that we should be able to experience monocular stereopsis by viewing a single static 2D image with one eye closed. This is the concern of a second strand in the literature that started in the mid-1990s with the exploration of so-called paradoxical monocular stereoscopy by Koenderink et al. (1994) (see also Enright 1989, and Eby and Braunstein 1995); and has most recently been explored by Koenderink et al. (2013), Vishwanath and Hibbard (2013), Volcic et al. (2014), Vishwanath (2016), and Wijntjes et al. (2016). This literature also revived a long forgotten tradition of monocular stereoscopy by some of the finest minds in early-to-mid twentieth century vision science: von Rohr (1903), Claparède (1904), Holt (1904), Münsterberg (1904), Ames (1925a, b), Carr (1935), Schlosberg (1941), Gibson (1947), and Gabor (1960).

It is tempting to suggest that this early literature was a casualty of Julesz (1960), and the subsequent conflation of stereopsis with binocular disparity. But in truth monocular stereopsis had already been rejected by Ogle (1959) on purely experiential grounds. Ogle suggested that Wheatstone's (1838) stereoscopic line drawings demonstrated the 'fundamental' difference between binocular depth *perception* on the one hand, and the mere *conception* of depth available from monocular viewing: he argued that there was absolutely no impression of stereoscopic depth from these simple line drawings when they were viewed monocularly, and yet when they were viewed in a stereoscope such images produced a vivid impression of depth. This is true, but Ogle's mistake was to generalise from this example to all instances of monocular depth: he used this example to conclude that stereoscopic depth perception was the

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 9/34

'single outstanding function of vision with the two eyes', that was 'not even suggested by vision with one eye alone.'

According to the contemporary literature, Ogle's mistake was to rely on simple line drawings as being representative of all 2D images: it would argue that once depth cues such as perspective and shading are added to the image, there is ample evidence that (a) monocularly viewed images produce a depth percept, and also (b) that synoptic viewing (viewing two identical 2D images in a stereoscope) can significantly accentuate this impression of depth: see Koenderink et al. (1994) and Wijntjes et al. (2016). On the one hand, in Chap. 3 I question whether the evidence in favour of monocular depth perception from 2D images really goes to our *perception* rather than our *cognition* of depth. But on the other hand, I also resist Ogle's suggestion that if we fail to perceive depth in a monocularly viewed 2D image, this necessarily implies that monocular vision of the 3D world must also lack depth.

3. Monocular Stereopsis in a 3D World: But how might we explain the monocular perception of depth in the 3D world if it is absent in a 2D image? Well, objects distributed throughout space in a 3D world will be subject to various different degrees of defocus blur. Traditionally when defocus blur has been treated as a depth cue, it has been regarded as just another pictorial cue alongside perspective and shading. But for this pictorial account of defocus blur to work it has to penetrate our subjective visual experience. Consequently, since defocus blur is typically apparent less than 4% of the time, it is commonly assumed that defocus blur must be a depth cue with only limited application: see Sprague et al. (2016). By contrast, in Chap. 4 I argue that if my contention in Chap. 3 is correct, and we do not perceive depth from perspective or shading, then we need another explanation for why we appear to be able to see depth when we look at the 3D world with one eye. The only solution, I suggest, is that just as the visual system can rely on sub-threshold defocus blur in order to guide accommodation, it can also rely on sub-threshold defocus blur in order to determine, in a very rough sense, the depth relations in the scene.

Now whilst sub-threshold defocus blur might account for the perceived geometry of a monocularly viewed scene, what about its scale?

A common assumption is that we can scale a monocular scene using accommodation (the tension in the ciliary muscles that control the intraocular lens indicating the distance of the focal plane). But as I argue in Chap. 4, there are both theoretical and empirical considerations that militate against this hypothesis. Instead, I conclude that we should be open to the idea that monocular vision does not convey scale, and that scale is only something that we *cognitively impute* to the scene. Indeed, Chap. 4 raises the prospect that this might hold true for binocular vision as well.

4. Extracting Depth from Binocular Disparity: But what about the claim that started the anti-Cognitive Revolution in the first place, namely Ogle's insistence that figure-ground separation was a prerequisite for extracting depth from binocular disparity? Well, so far as this question is concerned, if the 1960s had marked a 'retreat from mentalism', then the neo-Gestalt Revolution of the 1980s (see Ramachandran 2006) marked a return: although Julesz (1960) was right that figure-ground separation was not a *prerequisite* for extracting depth from disparity, Ramachandran and Cavanagh (1985) demonstrate that the subjective contours of Kanizsa figures (see Fig. 7) appear to influence to the structure of the depth that is derived from binocular disparity (see also Ramachandran 1986; Nakayama et al. 1989; and Mather 1989, although Mather leaves it open whether subjective contours really influence the extraction of depth from disparity, or are themselves merely a consequence of it).

Furthermore, Zhou et al. (2000) argue that V2 (the secondary visual cortex) not only identifies contours but also differentiates between figure and ground, whilst Qiu and von der Heydt (2005) go one step further by suggesting that V2 achieves this differentiation between figure and ground by employing disparity and Gestalt rules alongside one another (see also Nakayama 2005; Ramachandran and Rogers-Ramachandran 2009; and von der Heydt 2015).

The suggestion that *meaning* can be attributed to monocularly viewed images by the visual system, and that this meaning can be used to disambiguate the signal from binocular disparity, is an intriguing one; but if my analysis of monocular vision in the previous subsection is correct, its importance is liable to be overstated:

First, we would have to ensure that Mather's (1989) alternative explanation for this phenomenon had been entirely ruled out. It is plausible that the subjective contours are experienced in stereoscopic depth not because (a) figure and ground have been interpreted in the monocular image before the extraction of depth from disparity, but because (b) this is the most parsimonious solution as to how the sparse stereo elements fit together coherently in a 3D scene.

Second, even if it turns out that figure-ground separation is relied upon to disambiguate stereograms with sparse disparity information, such disambiguation is *2D-plus* rather than *3D*: it involves (a) the 2D *segmentation* of the image, followed by (b) the *ordering* of its layers, but this is still far removed from (c) the attribution of *depth* to these layers. In this sense, figure-ground cues are more like the recognition of words on a page than the attribution of depth: words are recognised as being *on* the page, even though there is *no depth* between the words and the page.

Third, to the extent that the literature on disambiguating binocular disparity claims anything more, and moves from (a) using monocular cues to *disambiguate* binocular disparity to (b) placing monocular cues *in conflict* with binocular disparity (see Qiu and von der Heydt 2005), then it strays into the cue-conflict literature which is fully explored in Chap. 2 (this is the reading of Qiu and von der Heydt 2005 advanced by Burge et al. 2010). Indeed, we might use the cue-conflict literature to try to better understand which side of the perception–cognition divide V2 lies (when conjoined with the results in Qiu and von der Heydt 2005, my position in Chap. 2 logically entails that V2 is engaged in cognition rather than perception).

Finally, even if the visual system relies upon subjective contours to extract depth when faced with (a) two flat 2D images, and (b) sparse disparity information by which reconcile these two images, I am sceptical that Ramachandran and Cavanagh's (1985) findings have any general application outside this context. Under my account, the real world already provides the visual system with an optical cue to figure-ground relations: defocus blur. So stereoscopic viewing of 2D images with sparse disparity information begins to look like an artificial, contrived, and arguably misleading, basis upon which to understand the relationship between stereopsis and pictorial processing.

Indeed, as Mach recognised over 140 years ago, if you already have a monocular conception of stereopsis, then binocular stereopsis begins to resemble a secondary process that merely accentuates the prior monocular processing (see Banks 2001). Similarly, Koenderink et al. (2015b) describe monocular stereopsis as 'stereopsis proper' and suggest that binocular stereopsis is at least partly, but probably largely, to be explained in monocular terms. Such a conclusion would also make sense from an evolutionary perspective: monocular stereopsis must have emerged in herbivores before binocular stereopsis emerged in predators. Consequently, we would expect the binocular depth processing that emerged to be parasitic upon the monocular depth processing that already existed.

3 Visual Cognition

But even if I am right and pictorial cues do not contribute to our perception of depth, the pictorial cues in cue-conflict stimuli (Chap. 2) and 2D images (Chap. 3) clearly contribute to *something*: if it is not our *perception* of depth, then what is it? I would argue that they contribute to an *automatic* (i.e. not consciously or deliberately made, and often involuntary) *post*-perceptual *evaluation* (or *judgement*) of the scene. Under this account pictorial cues are not *perceptual* cues but *cognitive* cues. But, and this is the important point, they are cues to a relatively self-contained module of *cognition*, divorced from *conscious deliberation*.

In this sense, there is an affinity between my position and Cavanagh's (2011) account of *visual cognition* as an *unconscious* (we are unaware of it at work), *automatic* (we do not have to do anything), and *involuntary* (we often cannot overrule it) process that attributes *meaning* to sensory data *before* conscious deliberation. But the key difference is that for Cavanagh, visual cognition operates as at the level of perception (Fig. 2).

Cavanagh (2011) documents the 'extraordinarily sophisticated' perceptual inferences of visual cognition that are distinct from conscious deliberation. The classic example is the Müller-Lyer illusion (see Chap. 2, Fig. 12): the illusion still persists even though we *know* that the lines in the Müller-Lyer illusion are the same length. But whilst Pylyshyn (1999) interprets the persistence of the Müller-Lyer illusion as evidence of *cognitive impenetrability* (i.e. of perception being immune from cognitive influence), Cavanagh insists that it actually evidence of *cognitive independence*:

> Pylyshyn calls this cognitive impenetrability but we might see it as cognitive independence: having an independent, intelligent agent—vision—with its own inferential mechanisms.

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 13/34

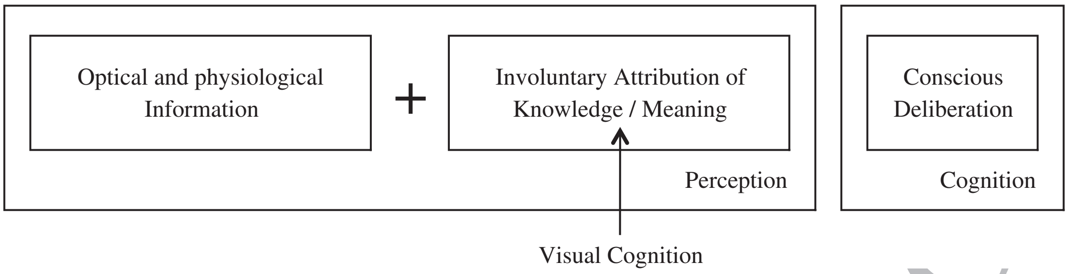

Fig. 2 Visual cognition according to Cavanagh (2011)

For Cavanagh, the inferential mechanisms of visual cognition play an essential role in determining the content of perception. Whether they are the top-down or high-level inferences of Bayesian Cue Integration, or merely the mid-level inferences associated with Gestalt Psychology, the point is the same: the retinal information is insufficient to specify the percept, so inferential mechanisms are relied upon to determine which percept out of the many possible percepts we in fact see. This is not to constrain the form these inferences must take: as Cavanagh observes, they may be based on likelihood, bias, or even a whim. But the important point is that whatever form these inferences take, the visual system uses them to reject the many possible alternatives that were just as consistent with the raw sensory data as the eventual percept.

Nor is Cavanagh's account in tension with Firestone and Scholl's (2016a, b) recent work on cognitive impenetrability. Whilst Firestone & Scholl are refreshingly robust about the need to distinguish between *perception* and *cognition*, by *cognition* they mean the thoughts, desires, and emotions of the New Look literature (see Bruner and Goodman 1947). They are motivated by the '*revolutionary* possibility' that what we see is directly influenced by what we think, want, and feel. By contrast, Firestone & Scholl explicitly exclude the unconscious inferences that Cavanagh has in mind as being in any way controversial or suggestive of cognitive penetration, claiming that a good litmus test is whether such inferential processes continue to operate reflexively in spite of our own conscious deliberations (as is the case with the vast array of visual illusions, from the Müller-Lyer illusion to the hollow-face illusion and Reverspectives).

Now whilst I agree with Cavanagh (2011) that there appears to be a relatively self-contained module of visual cognition, I would argue (at least in so far as depth perception is concerned) that this module ought

Layout: **A5 HuSSci** Book ID: **455577_1_En** Book ISBN: **978-3-319-66293-0**
Chapter No.: **1** Date: **11 November 2017 14:31** Page: **14/34**

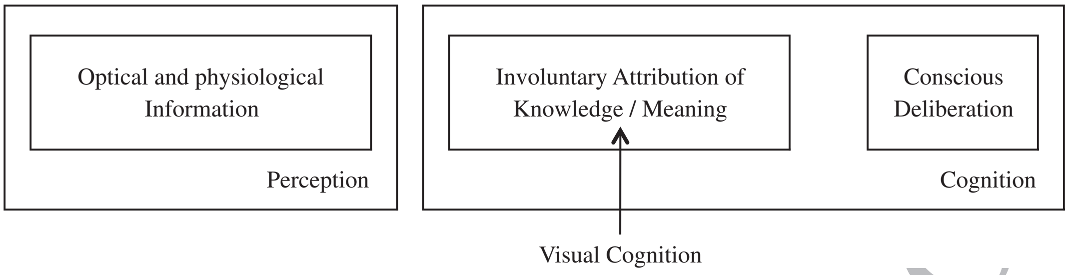

Fig. 3 Visual cognition according to my alternative account

to be regarded as *post-perceptual*, since it does not appear to affect the actually perceived geometry of the scene, but only our *judgement* or *evaluation* of it (Fig. 3).

A useful analogy I explore in Chap. 3 is with *reading*: being able to understand the meaning of a word doesn't change the perceptual appearance of the text, but nor does it rely upon conscious deliberation. Instead, it appears to be both a *post-perceptual* and an *automatic* and *involuntary* process of attributing meaning to what we see. Similarly, I would argue that something 'looking flat', 'looking round', 'looking square', or 'looking symmetrical' is not really a *perceptual* claim, but a *post-perceptual* attribution of depth or shape meaning: a *judgement* or *evaluation* about what we see. And I would argue that pictorial cues *bias* our evaluation of depth, rather than informing our perception of depth. On the other hand I agree with Cavanagh that visual cognition must be *pre-deliberative*, since it biases our evaluation of depth in a way that is apparently *not* open to rational revision (in this sense *visual cognition* is not only *automatic*, but also *involuntary*). Indeed, often the only way to counteract these biases is to introduce a visual comparator (see Chap. 2); in a sense, to change the cognitive task from an *evaluation* to a simple *comparison*.

This debate is not only important for depth perception, but also the wider question of the role of cognition in vision. After all, depth from pictorial cues represents the thin edge of a very significant cognitive wedge for Cavanagh. And, as more and more complex phenomena (such as *causation* and *intentionality*) are attributed to vision, the more intelligent Cavanagh insists the visual system must be:

> …the unconscious inferences of the visual system may include models of goals of others as well as some version of the rules of physics. If a 'Theory

Layout: A5 HuSSci | Book ID: 455577_1_En | Book ISBN: 978-3-319-66293-0
Chapter No.: 1 | Date: 11 November 2017 14:31 | Page: 15/34

of Mind' could be shown to be independently resident in the visual system, it would be a sign that our visual systems, on their own, rank with the most advanced species in cognitive evolution.

By contrast, one of the virtues of my account is that we do not have to posit the existence of 'an independent, intelligent agent—vision' to explain these increasingly complex phenomena; instead, we simply recognise that post-perceptual human cognition may be broken into relatively independent modules.

4 Four Conceptions of Meaning

In this final section, I outline four of the leading accounts of stereopsis: Pictorial Cues, Cue Integration, Gestalt Psychology, and Intentionality, and explore the kinds of meaning that each account suggests must be attributed to the raw sensory data before content can be extracted from it and/or attributed to it:

1. Pictorial Cues: I will explore perspective and shading as a means by which to understand the extraction of depth from pictorial cues more generally:

(a) Perspective: In 1903, Moritz von Rohr developed 'The Verant, a New Instrument for Viewing Photographs from the Correct Standpoint' for Carl Zeiss based upon the work of Allvar Gullstrand (see von Rohr 1903). This monocular lens ensured that observers could view a 2D image from its centre of projection, and for von Rohr this (in addition to setting accommodation at infinity) explained the impression of monocular depth that subjects reported: by placing their eye at the centre of projection, the subject experienced the very same perspective cues that they would have experienced had their eye been placed at the entrance pupil of the camera.

This claim was explored by Holt (1904) and Schlosberg (1941). As a disciple of Holt, and a close associate of Schlosberg's, Gibson (1947) could not ignore the implications of this observation. In his work for the US military during WWII, he agreed that if a single static 2D image was viewed monocularly, whilst eradicating any cues to flatness, the observer was liable to experience a monocular impression of visual depth equivalent to binocular stereoscopic viewing. Indeed, Gibson (1947) drew the

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 16/34

conclusion that if binocular disparity appeared to contribute little to our impression of depth from 2D stereoscopic images, then it must also contribute little to our impression of depth from the 3D world, and this led Gibson (1950) to embrace an account of depth perception according to which binocular disparity played a largely insignificant role.

But we still need to explain why a static 2D image viewed from its centre of projection should induce an impression of depth? For von Rohr the answer was clear: viewed in this way, the observer experiences the very same perspective cues they would experience had they been present in the real world scene. But as Gibson observed, this explanation only poses a further problem: namely, why should perspective cues *from a real world scene* give rise to a monocular impression of depth in the first place? Gibson toyed with this question for much of his 50-year career, although the emphasis appears to shift away from monocular stereopsis towards pictorial depth: for instance, Schlosberg (1941) is cited in Gibson (1966) but not Gibson (1971) or Gibson (1979). One gets the impression that Gibson never fully resolved this question to his satisfaction. As he recounted just before his death (in Gibson 1979), he repeatedly revised his theory of pictorial cues, leaving a catalogue of abandoned accounts: Gibson (1954, 1960), and Chap. 11 of Gibson (1966). The intractable problem for Gibson (1979) was that perspective is *indeterminate*: it might specify *some* invariant features the scene must have, but it is neutral as between the various competing arrangements that satisfy these features. Indeed, this realisation led Gibson (1979) to ultimately reject monocular stereopsis from a 2D image, an insight that had previously meant so much to him:

> The purveyors of this doctrine disregard certain facts. The deception is possible only for a single eye at a fixed point of observation with a constricted field of view… This is not genuine vision, not as conceived in this book.

And yet for contemporary neo-Gibsonians, Gibson's most difficult case turns out to be their easiest. Consider Rogers and Gyani's (2010) discussion of Patrick Hughes' 'Reverspectives', *protruding* physical forms that are painted as if they are *receding* in perspective (in this instance, the canals of Venice) (Fig. 4).

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 17/34

Fig. 4 Patrick Hughes in his studio. © Patrick Hughes. For more information please see: http://www.patrickhughes.co.uk/

Rogers and Gyani (2010) suggest that when stationary observers view this artwork monocularly, they perceive it as a scene *receding* in depth rather than its actual physical form (i.e. as an object *protruding* in depth). For Rogers and Gyani, the reason for this depth inversion is 'obvious': 'What we see is consistent with the information provided by the perspective gradients'. But the question is not whether the illusory percept is *consistent* with perspective. That is a given. Instead, the question is why *this* percept is chosen out of the innumerable consistent possible interpretations? This was Gibson's question. And for Rogers and Gyani, the answer is that the illusory percept is not just *consistent* with perspective, but *specified* by it:

> …we should not be surprised that we see 'reversed' depth when these delightful artworks are viewed monocularly *because this is what the perspective information is telling us…* (emphasis added)

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 18/34

By contrast, I would argue that there is no such thing as *perspective information*, only *optical information* to which *perspective meaning* has been attributed.

Indeed, *perspective meaning* is something that has to be learnt. This is demonstrated by the fact that perspective images mean nothing to those with newly restored sight if they have been blind all their lives. For instance, when Sidney Bradford had his sight restored (see Gregory and Wallace 1963) he was immediately able to understand capital letters and clock-faces (as they had been taught to him via touch) but not, as Gregory (2004) explains, pictures: pictures looked flat and meaningless to him, in spite of the fact that he could judge the size and distance of objects that were already familiar from touch (e.g. chairs scattered around the ward). Furthermore, the process of learning to attribute meaning to perspective is gradual: even after six months Mike May, another formerly blind patient, was unable to identify wireframe drawings of cubes in any orientation, describing them as 'a square with lines' (see Fine et al. 2003).

So whilst Rogers and Gyani (2010) may dismiss the experience of a static monocular observer of a Reverspective as uninformative (at one point suggesting that it 'cannot tell us anything about the visual system'), I would argue that it would, in fact, tell us something very significant: namely that (if Rogers and Gyani are correct) the visual system utilises a learnt form of *meaning*, perspective meaning, to determine the content of stereopsis. You might object that calling this *meaning* puts the point too finely. After all, Rogers and Gyani are keen to emphasise the low-level nature of perspective: they demonstrate that a simple wireframe Reverspective can be just as effective as a fully rendered scene, and suggest that converging line junctions are sufficient to induce a depth percept all by themselves.

Similarly Cavanagh (2011) excludes low-level processes from visual cognition. Indeed, there is a similarity between the way subjective completion can lead to perverse results (for instance, the conjoining of the front of one animal with the back of another into an impossibly long form), and the way in which the visual system is liable to come to an automatic interpretation of perspective cues, even if it is obviously wrong (for instance, in the context of an Ames Rooms). In the context of subjective completion, Cavanagh asks: 'Given this very lawful behaviour, we might ask if there is anything really inferential here'. The same, Rogers

Layout: A5 HusSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 19/34

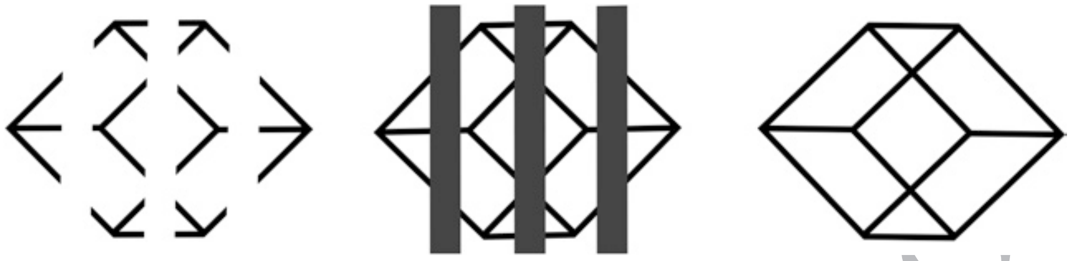

Fig. 5 Isolated line junctions inspired by Nakayama (1999)

and Gyani would argue, could be asked in the context of extracting depth from perspective.

But I think it would be a mistake to draw a distinction between *lawful* processing on the one hand and *cognition* on the other. Cavanagh's account is liable to run three distinct concerns together: (a) *inference* (incorporating some notion of *problem-solving*), (b) *complexity* (incorporating some notion of *intelligence*), and (c) *choice* (incorporating some notion of *agency*). By contrast, I think *agency* is unhelpful in this context: logic, mathematics, and linguistics are all forms of rule-based reasoning that clearly ought to qualify as cognition if the visual system is engaged in them. Nor can we dismiss the rule-based extraction of depth from perspective as *just processing*, since everything in the brain is ultimately 'just' rule-based processing. Indeed, this is reflected in Cavanagh's own description of inferences:

> Note that an inference is not a guess. It is a rule-based extension from partial data to the most appropriate solution.

So the choice is between *simple* rule-based processing and *complex* rule-based processing; Rogers and Gyani (2010) may be an instance of the former, and Cavanagh (2011) primarily concerned with the latter, but the *complexity* of the meaning being attributed makes no difference to my account: rudimentary meaning is still meaning.

But my second response to Rogers and Gyani is to question just how rudimentary the extraction of depth from perspective really is? Whilst Rogers and Gyani suggest that the visual system exploits line junctions, Nakayama (1999) demonstrates that line junctions all by themselves are not necessarily that informative (Fig. 5).

Layout: **A5 HuSSci** Book ID: **455577_1_En** Book ISBN: **978-3-319-66293-0**
Chapter No.: **1** Date: **11 November 2017 14:31** Page: **20/34**

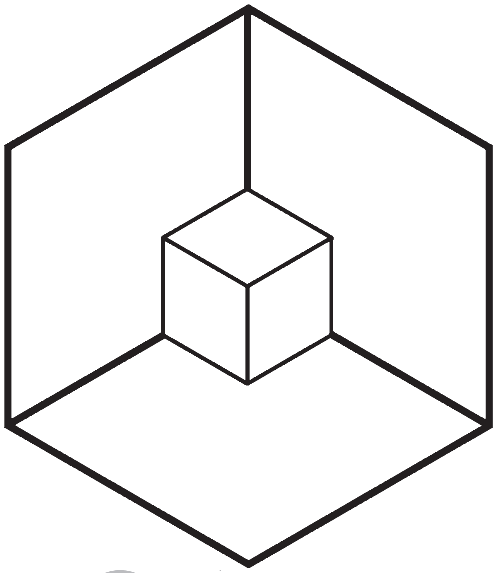

Fig. 6 Tristable perspective figure inspired by Poston and Stewart (1978) and Wallis and Ringelhan (2013)

First, these junctions can easily be given a 2D interpretation. Second, even if they are given a 3D interpretation, it is far from obvious that they represent three angles of equal size. Instead, this interpretation only appears to emerge once the individual junctions are themselves seen as part of a coherent whole: it is as if the eight junctions become eight simultaneous equations, to which 90° is the only rational solution. But if this is the case, and the *whole* specifies the *parts*, then this is far from a *low-level* process. Indeed, we reach the same conclusion by considering multi-stable cubic volumes whose components are liable to be interpreted as a coherent whole (either as a small cube against a background or as large cube with a small chunk taken out of it) even though a small perspective cube in front of a large perspective cube is just as permissible an interpretation (Fig. 6).

But the deeper concern is that by focusing on a ‘carpentered world’ of parallel lines and right-angles (such as cubic volumes, Ames Rooms, and Reverspectives), we risk massively underestimating the complexity of the processes that extract depth from perspective. It is easy to forget that the visual system did not evolve in response to a ‘carpentered world’, and that the forms it did evolve in response to were positively irregular by comparison. Consequently, the visual system’s response to perspective cues is likely to be much more nuanced than its automatic interpretation of cubic volumes would suggest. Indeed, we do not need to appeal to the positively irregular forms of human evolution to illustrate this

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 21/34

point: as Knill (2007) demonstrates, even extracting perspective information from a regular shape like an ellipse depends heavily on prior knowledge and/or assumptions about the most likely interpretation of a given scene. Landy et al. (2011) explain the kind of complex scene statistics that would have to be employed by subjects:

> The generative model for the aspect ratio of an ellipse in the image depends on both the 3D slant of a surface and the aspect ratio of the ellipse in the world. The aspect ratio of the ellipse in the world is a hidden variable and must be integrated out to derive the likelihood of slant. The prior distribution on ellipse aspect ratios plays a critical role here. The true prior is a mixture of distributions, each corresponding to different categories of shapes in the world.

Furthermore, according to Knill (2007) the visual system doesn't just engage in natural scene statistics, it also engages in real-time *perceptual learning*: The subjects in Knill (2007) initially assumed that any ellipse in the visual field must be a slanted circle. But as the experiment progressed they encountered a number of patently non-circular ellipses, and so learnt that the circularity of the ellipses could not be assumed. Consequently, when the slightly non-circular ellipses that had previously been judged as circular earlier in the experiment were shown again, the subjects now correctly identified them as non-circular.

b. Shading: So extracting depth from perspective proves to be anything but a low-level process, and the same appears to be true for shading:

First, shading is a change in the *luminance* of a surface, but our interpretation of surface luminance is a complex phenomenon, that is only partly determined by the amount of light that is reflected from the object to the retina: see Gilchrist (2006).

Second, extracting depth information from *changes* in luminance requires a mechanism that can take those changes into account. But Tyler (2006) suggests that such a mechanism would have to be surprisingly complex; certainly well beyond the range of early visual image filters which Morgan and Watt (1982) have estimated only extends to 2–3 arc min (1/30th to 1/20th of a degree). So if shape is interpolated on the basis of changes in luminance over the surface of the object, a mid-level or higher-level process must be responsible.

Third, accurately extracting depth information from shading requires prior knowledge about the direction of illumination (see Pentland 1982). For Wagemans et al. (2010), this is evidence that 'the shading cue is inherently ambiguous', leading them to give up on inverse optics in Koenderink et al. (2015a) and instead treat shading merely as an instance of 'relief articulation', much like contour-lines drawn on a map to convey relief. The only alternative is to appeal to a default assumption about the illumination in the scene. Three candidates have been advanced: The first is to posit a single strong *overhead* light-source (i.e. the sun): see Ramachandran (1988). The second is to suggest that light, having been reflected between the atmosphere and the ground multiple times, is *diffuse*: see Gibson (1979) and Chen and Tyler (2015). The third is to adopt an *ecological* perspective, according to which both are permissible: overhead light on a sunny day, and diffuse light on a cloudy day; but this entails an even more sophisticated process of extracting depth from shading given that overhead and diffuse light cast such very different shadows, see Langer and Bülthoff (2000).

Fourth, once we have finally settled on an appropriate assumption, we still have to use it to extract the relevant depth information from shading, and this promises to be another complex undertaking: Tyler (1998) and Chen and Tyler (2015) have argued that under the diffuse light assumption the visual system can adopt a quick and easy 'dark is deep' rule of thumb, but Langer and Bülthoff (2000) and Todd et al. (2015) have demonstrated that even under diffuse light dark does not necessarily mean deep, and so have questioned the ecological validity of this approach.

2. Cue Integration: As these discussions illustrate, the process of extracting depth information from a single depth cue such as perspective or shading implies a significant degree of complexity. Consequently, individual depth cues are liable to provide us with only partial, noisy, or contradictory depth information. But if this is the case then a second stage of cognitive processing is required in order to integrate and reconcile these various contradictory sources of depth information into a single coherent percept. And prior knowledge is thought to play a central role in this integration process.

In the contemporary literature this reliance on prior knowledge is typically articulated in Bayesian terms, and there is no doubt that the

Layout: A5 HuSSci | Book ID: 455577_1_En | Book ISBN: 978-3-319-66293-0
Chapter No.: 1 | Date: 11 November 2017 14:31 | Page: 23/34

Bayesian literature of the last couple decades has brought greater statistical sophistication to bear on this question. Nonetheless, as Trommershäuser et al. (2011) observe, the fundamental principle that underpins Cue Integration was already apparent in Helmholtz's (1866) unconscious inferences, and even in the work of al-Haytham (c.1028–38). Similarly, Seydell et al. (2011) suggest that we might regard Cue Integration as the *veridical* counterpart to Gregory's (1970) hollow-face *illusion*: whilst the visual system's reliance upon prior knowledge may give rise to illusions in certain artificially contrived contexts (e.g. the misinterpretation of a hollow mask), ordinarily a reliance on prior knowledge only improves the visual system's ability to estimate the true state of the world.

For Cue Integration this reliance on prior knowledge is a *prerequisite* for perception. This is sometimes overlooked in the literature, where there can be a tendency to pit 'top-down' prior knowledge against 'bottom-up' sensory data. For instance, Nguyen et al. (2016) suggest that what we see depends upon two types of influences that can be in competition: (a) 'bottom-up' cues such as edge orientation, the direction and speed of motion, luminance and chromatic contrast, and binocular disparity, and (b) 'top-down' influences such as endogenous attention, expectations, and stored visual knowledge, of which they advance Bayesian Cue Integration as an example. But to suggest that 'top-down' processing is either in conflict with, or merely influences, the 'bottom-up' sensory data is to underestimate the importance of 'top-down' processing for Cue Integration accounts: according to Cue Integration 'top-down' processing is the *only* way of attributing depth meaning to sensory data, without which the sensory data would simply have no content. So it is not as if 'top-down' processing merely *influences* or *competes with* the 'bottom-up' sensory data, or that if the 'top-down' processing were absent 'bottom-up' sensory data would be free to determine the percept; instead, 'top-down' processing constitutes perception under a Cue Integration account.

Finally, although Trommershäuser et al. (2011) observe that 'Bayesian statistics is emerging as a common language in which cue-combination problems can be expressed', this is not the only articulation of Cue Integration in the literature. Indeed, since the late 1990s Domini, Caudek, and colleagues have emphasised the *non-veridical* and often *mutually inconsistent* nature of Cue Integration (see Domini and

Caudek 2011). Especially important for Domini and Caudek (2011), as well as for Scarfe and Hibbard (2011), is the possibility that individual cues might be *biased*. Domini and Caudek argue that if it can be demonstrated that bias really is pervasive in the visual system, then this should have a transformative effect on how we ought to conceive of vision: Is the goal of vision to recover a veridical depth map of the scene? Or is it, as Domini and Caudek suggest, merely concerned with ensuring that we can effectively interact with the environment?

Indeed, this concern with successful interaction, rather than recovering a metric depth map, reflects a recent trend in cognitive science which Engel et al. (2013) term *the pragmatic turn*. As Engel et al. (2016) explain:

> Cognitive science is witnessing a pragmatic turn away from the traditional representation-centered framework of cognition towards one that focuses on understanding cognition as being 'enactive.' The enactive view holds that cognition does not produce models of the world but rather subserves action…

But even those who continue to articulate Cue Integration in *representational* terms are liable to (a) question the wisdom of associating Cue Integration with *optimality*: see Rosas and Wichmann (2011), or (b) suggest a *less formulaic* approach, according to which Cue Integration is closer to testing a hypothesis (see Gregory 1980 and Tyler 2004) or playing 20 questions with nature (see Kosslyn 2006 and Cavanagh 2011).

3. Gestalt Psychology: 'Gestalt' is German for 'pattern' or 'shape', although 'configuration' is closer to what was intended (see Rock and Palmer 1990), with the central argument of Wertheimer's (1924) principle of 'holism' being that we directly perceive *configurations* or *integrated wholes* whose properties are greater than the sum of their parts. The classic illustration of this is the Kanizsa Triangle (Fig. 7): The unavoidable impression is of a white triangle occluding three black circles and a wireframe triangle. But no white triangle is specified by the stimulus. Nor are the black circles or the wireframe triangle. And there is no sense in which an inverse optical account that failed to specify these circles and triangles would be incomplete. So this is taken as evidence that something more than inverse optics must be going on. And the unresolved question of the last couple of decades is what this something

Layout: A5 HusSci | Book ID: 455577_1_En | Book ISBN: 978-3-319-66293-0
Chapter No.: 1 | Date: 11 November 2017 14:31 | Page: 25/34

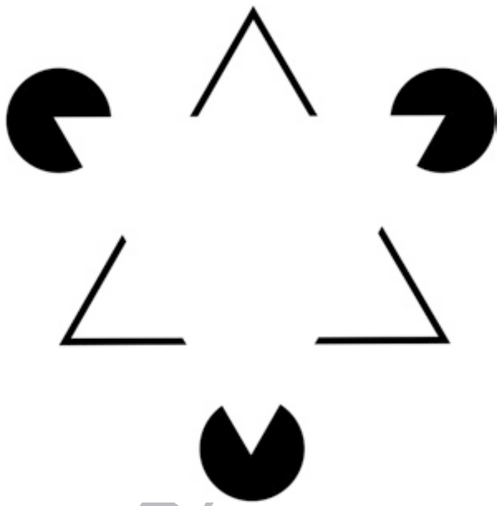

Fig. 7 Kanizsa triangle inspired by Kanizsa (1955)

more is, and how exactly it relates to Cue Integration? Specifically, are Gestalt phenomena such as the Kanizsa Triangle (a) *an alternative to*, (b) *supplementary to*, or (c) simply *just an application of* Cue Integration?

As Wagemans et al. (2012a) observe, most textbooks will contain a chapter on Gestalt phenomena but leave their relationship with the rest of the literature ambiguous. But in another sense it is no surprise that this tension between these Gestalt phenomena and Cue Integration has not been resolved because we are still unsure as to what exactly is driving the Gestalt phenomena in the first place: is it *likelihood* or is it *simplicity*? As Wagemans et al. (2012a, b) ask, do we see the white triangle in Fig. 7 because it is the *most likely* interpretation of the stimulus, or merely because it is the *most straightforward* one?

If it is the former, then Gestalt principles are subsumed under Cue Integration. And certainly, Wagemans et al. (2012a) would not shy away from this conclusion, suggesting that groupings could be based on probabilistic models derived from natural scene statistics. Indeed, even those who embrace the alternative principle of *Prägnanz* (or *simplicity*) are often inspired to do so by Structural Information Theory on the basis that in absence of knowledge about the environment the simplest solution is often the most likely: Wagemans et al. (2012a) suggest that evolution may well have built a surrogate for *likelihood* into the visual system via *simplicity*.

A commitment to *Prägnanz* (or *simplicity*) is perhaps the closest to the classical view of Gestalt as employing innate laws of perceptual organisation.

Layout: **A5 HuSSci** Book ID: **455577_1_En** Book ISBN: **978-3-319-66293-0**
Chapter No.: **1** Date: **11 November 2017 14:31** Page: **26/34**

But even here, theoretical abstraction has to give way to empirical reality. As Wagemans et al. (2012a) observe, Gestalt principles are no longer thought of as simply pre-attentive grouping principles, but operate instead at multiple levels and can be heavily influenced by past experience.

By contrast, a third interpretation of Gestalt brings its grouping principles closer to Intentionality. Koenderink (2010) suggests that 'perceptual organisation' is a process of attributing *subjective meaning* to a scene; so rather than asking which is the *statistically most likely* interpretation, or even the *simplest* one, we ask which is the *most rational*: 'There is simply no way to "transform" mere structure into meaning, you—as *perceiver*—have to *supply* it.'

4. Intentionality: Indeed Albertazzi et al. (2010) argue that the insights of early-twentieth century Gestalt Psychology derive from a deeper truth articulated in the late-nineteenth century by Brentano (1874), namely the *act of intentional reference*, according to which:

> …the structure of a process of seeing, thinking, judging, and so on is that of a *dynamic whole endowed with parts* in which the parts are nonindependent items, and that this act can give rise to relatively different outputs based on *subjective completions*…

But Albertazzi et al.'s aspirations for Intentionality go further still:

> The linking theme is the foundational role of perception as the origin of every potential level of signification, from the most concrete to the most abstract (Arnheim 1969), and a particularly strong interest in the *qualitative* aspects of experience, for within these lie the clues to a richer semantic theory of information.

Albertazzi et al. illustrate their point that vision ought to be understood as much in terms of *qualities* as the *quantities* of geometry and scale, with the example of aesthetic properties: they argue that we *see* aesthetic properties, and yet there is no place for aesthetic properties amongst the traditional primary (geometry and scale) and secondary (colour) qualities of vision. Albertazzi et al. argue that what is required to accommodate such properties is, instead, 'a theory of perception that sees qualitative phenomena and the subjective operations of the observer as foundation.'

But why is this of any interest to us? After all, isn't stereopsis simply concerned with the *quantities* of scale and geometry that are already admitted? Not so, according to Vishwanath (2010), who argues that a 2D image can effectively convey the three-dimensional properties of the scene *without* stereopsis. We will explore in Chap. 3 whether this position is sustainable, but Vishwanath takes this as evidence that stereopsis must reflect *something more* than the three-dimensional properties of the scene, specifically a *quality* of vision that reflects a more *subjectively meaningful* layer of depth, namely 'the depth used to guide motor function' (Volcic et al. 2014). But what does this mean?

Well, to understand Vishwanath's account of *stereopsis* we first have to understand his account of the *surfaces* of objects. Vishwanath (2010) argues that the surfaces of objects ought to be understood as invitations to interact with the world; specifically, they are *anticipatory* structures: the presentation of complex motor *plans*. But how are we to test the validity of such plans? Will we successfully interact with the world if we follow them? Or will we fail? The obvious answer is to simply to try them and see: some motor plans will result in success, others in failure. But from an evolutionary perspective this has huge costs, with every failure being potentially fatal. And this is where a role for stereopsis as a subjective quality of visual experience begins to emerge for Vishwanath:

> Conveniently, my perceptual system has given me a way of being implicitly weary of putting all faith in the 3D presentation before me: by modulating the perceived plastic quality of that 3D presentation.

And so stereopsis becomes the means by which the visual system conveys the *reliability* of the complex motor plans that surfaces represent. Specifically, whilst Vishwanath suggests that our impression of the geometry of an object is largely accurate (even from a 2D image, where stereopsis is absent), what is required to successfully interact with this object is that we not only have (a) its geometry, but also (b) access to reliable egocentric distance information by which to scale the geometry: Is it a small object up close or a large object far away? Consequently, for Vishwanath, stereopsis is the visual system's way of communicating to the observer the precision with which it is able to scale the geometry of the scene or object. Whether this is a sustainable position is explored in Chap. 3.

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 28/34

References

Albertazzi, L., van Tonder, G. J., & Vishwanath, D. (2010). *Perception beyond inference: The informational content of visual processes.* Cambridge, MA: MIT Press.

al-Haytham. (c.1028–1038). Book of Optics. In A. M. Smith (trans. & ed.) (2001). *Alhacen's theory of visual perception, volume two: English translation, transactions of the american philosophical society* (Vol. 91, Part 5). Philadelphia: American Philosophical Society.

Ames, A., Jr. (1925a). The illusion of depth from single pictures. *Journal of the Optical Society of America, 10*(2), 137–148.

Ames, A., Jr. (1925b). Depth in pictorial art. *The Art Bulletin, 8*(1), 4–24.

Ames, A., Jr. (1951). Visual perception and the rotating trapezoidal window. *Psychological Monographs, 65*(7), 324.

Ames, A., Jr. (1955). *An interpretative manual: The nature of our perceptions, prehensions, and behavior.* For the Demonstrations in the Psychology Research Center, Princeton University. Princeton, NJ: Princeton University Press.

Arnheim, R. (1969). *Visual Thinking.* Berkeley and Los Angeles, CA: University of California Press.

Aschenbrenner, C. (1954). Problems in getting information into and out of air photographs. *Photogrammetric Engineering, 20*(3), 398–401.

Banks, E. C. (2001). Ernst Mach and the episode of the monocular depth sensations. *Journal of the History of the Behavioural Sciences, 37*(4), 327–348.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth perception. *Journal of Physiology, 193,* 327–342.

Barry, S. (2009). *Fixing my gaze: A scientist's journey into seeing in three dimensions.* New York: Basic Books.

Berkeley, G. (1709). *An essay towards a new theory of vision.* Dublin: Printed by Aaron Rhames, at the back of Dick's Coffee-House, for Jeremy Pepyat, bookseller in Skinner-Row.

Bishop, P. O., & Pettigrew, J. D. (1986). Neural mechanisms of binocular vision. *Vision Research, 26*(9), 1587–1600.

Brentano, F. (1874). *Psychology from an empirical standpoint* (A.C. Rancurello, D.B. Terrell, & L. McAlister Trans.) (1973). London: Routledge.

Bruner, J. S., & Goodman, C. C. (1947). Value and needs as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42,* 33–44.

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

Cajal, S. R. (1904). Textura del Sistema Nervioso del Hombre y los Vertebrados (N. Swanson & L. W. Swanson Trans.) (1995). In *Histology of the nervous system of man and vertebrates.* Oxford: Oxford University Press.

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 29/34

Carr, H. A. (1935). *An introduction to space perception*. New York: Longmans, Green, & Co.

Cavanagh, P. (2011). Visual cognition. *Vision Research, 51*(13), 1538–1551.

Chen, C. C., & Tyler, C. W. (2015). Shading beats binocular disparity in depth from luminance gradients: Evidence against a maximum likelihood principle for Cue combination. *PLoS ONE, 10*(8), e0132658.

Claparède, E. (1904). Stéréoscopie monoculaire paradoxale. *Annales d'Oculistique, 132,* 465–466.

Cumming, B. G., & Parker, A. J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *Journal of Neuroscience, 19,* 5602–5618.

Domini, F., & Caudek, C. (2011). Combining Image Signals before Three-Dimensional Reconstruction: The Intrinsic Constraint Model of Cue Integration. In J. Trommershäuser, K. Körding, & M. Landy (Eds.), *Sensory cue integration*. Oxford: Oxford University Press.

Eby, D. W., & Braunstein, M. L. (1995). The perceptual flattening of three-dimensional scenes enclosed by a frame. *Perception, 24*(9), 981–993.

Engel, A. K., Friston, K. J., & Kragic, D. (2016). *The pragmatic turn: Toward action-oriented views in cognitive science*. Cambridge, MA: MIT Press.

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Science*, *17*(5), 202–209.

Enright, J. T. (1989). *Paradoxical monocular stereopsis and perspective vergence*. NASA, Ames Research Center, Spatial Displays and Spatial Instruments, N90-22922.

Fine, I., Wade, A. R., Brewer, A. A., May, M. G., Goodman, D. F., Boynton, G. M., et al. (2003). Long-term deprivation affects visual perception and cortex. *Nature Neuroscience, 6*(9), 915–916.

Firestone, C., & Scholl, B. J. (2016a). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences, 39,* 1–19.

Firestone, C., & Scholl, B. J. (2016b). Seeing and thinking: Foundational issues and empirical horizons. *Behavioral and Brain Sciences, 39,* 53–67.

Gabor, D. (1960). Three-dimensional cinema. *New Scientist*, 14th July 1960, 141.

Gibson, J. J. (1947). *Motion picture testing and research*. Research Reports, Report No. 7, Army Air Forces Aviation Psychology Program.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Gibson, J. J. (1954). A Theory of pictorial perception. *Audio-Visual Communication Review, 1,* 3.

Gibson, J. J. (1960). Pictures, Perspective, and Perception. *Daedalus, 89,* 216.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1971). The information available in pictures. *Leonardo, 4,* 27–35.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gilchrist, A. (2006). *Seeing black and white.* Oxford: Oxford University Press.

Gombrich, E. H. (1960). *Art and illusion: A study in the psychology of pictorial representation.* London: Phaidon.

Gregory, R. L. (1966). *Eye and brain: The psychology of seeing.* London: Weidenfeld & Nicolson.

Gregory, R. L. (1970). *The intelligent eye.* London: Weidenfeld & Nicolson.

Gregory, R. L. (1980). Perception as hypothesis. *Philosophical Transactions of the Royal Society B, 290*(1038),181–97.

Gregory, R. L. (2004). The blind leading the sighted. *Nature, 430,* 1.

Gregory, R. L., & Wallace, J. G. (1963). *Recovery from early blindness: A case study.* Experimental Psychology Society Monograph No. 2. Cambridge: Heffer.

Helmholtz, H. von. (1866). physiological Optics, Vol.3. In J. P. C. Southall (Trans. & ed.) (1925). *Treatise on physiological Optics.* New York: Dover.

Hering, E. (1865). Ueber stereoskopisches Sehen. *Verhandlungen des naturhistorisch-medizinischen Vereins zu Heidelberg, 3*, 8–11.

Hibbard, P. (2008). Can appearance be so deceptive? Representationalism and binocular vision. *Spatial Vision, 21*(6), 549–559.

Holt, E. (1904). Die von M. von Rohr gegebene Theorie des Veranten, eines Apparats zur Richtigen Betrachtung von Photographien by E. Wandersleb; The Verant, a New Instrument for Viewing Photographs from the Correct Standpoint by M. von Rohr; Der Verant, ein Apparat zum Betrachten von Photogrammen in Richtigen Abstande by A. Köhler. *The Journal of Philosophy, Psychology and Scientific Methods, 1*(20), 552–553.

Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell Labs Technical Journal, 39,* 1125–1162.

Kanizsa, G. (1955). Margini quasi-percettivi in campi con stimolazione omogenea. *Rivista di Psicologia, 49*(1), 7–30.

Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision, 7*(8), 13.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference.* Cambridge: Cambridge University Press.

Koenderink, J. J. (2010). Vision and information. In L. Albertazzi, J. van Tonder, & D. Vishwanath (Eds.), *Perception beyond inference: The informational content of visual processes.* Cambridge, MA: MIT Press.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1994). On so-called paradoxical monocular stereoscopy. *Perception, 23,* 583–594.

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 31/34

Koenderink, J. J., van Doorn, A., Albertazzi, L., & Wagemans, J. (2015a). Relief articulation techniques. *Art & Perception, 3*(2), 151–171.

Koenderink, J. J., van Doorn, A., & Wagemans, J. (2015b). Part and whole in pictorial relief. *i-Perception, 6*(6), 1–21.

Koenderink, J. J., Wijntjes, M. W. A., & van Doorn, A. J. (2013). Zograscopic viewing. *i-Perception, 4*(3), 192–206.

Kosslyn, S. M. (2006). You can play 20 questions with nature and win: Categorical versus coordinate spatial relations as a case study. *Neuropsychologia, 44*(9), 1519–1523.

Landy, M., Banks, M., & Knill, D. (2011). Ideal-observer models of cue integration. In J. Trommershäuser, K. Körding, & M. Landy (Eds.), *Sensory cue integration*. Oxford: Oxford 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.

Langer, M. S., & Bülthoff, H. H. (2000). Depth discrimination from shading under diffuse lighting. *Perception, 29*(6), 649–660.

Livingstone, M. (2002). *Vision and art: The biology of seeing*. New York: Abrams.

Mach, E. (1868). *Beobachtungen über monoculare Stereoscopie*. Sitzungsberichte der kaiserlichen Akademie, mathematische-naturwissenschaftliche Klasse, Wien, 58, 731–736.

Mach, E. (1886). *The analysis of sensations and the relation of the physical to the psychical* (C. M. Williams Trans.) (1959). New York: Dover.

Mather, G. (1989). The role of subjective contours in capture of stereopsis. *Vision Research, 29,* 143–146.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Science, 7*(3), 141–144.

Morgan, M. J., & Watt, R. J. (1982). Mechanisms of interpolation in human spatial vision. *Nature, 299,* 553–555.

Münsterberg, H. (1904). Perception of distance. *Journal of Philosophy, Psychology and Scientific Methods, 1*(23), 617–623.

Nakayama, K. (1999). Mid-level vision. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopaedia of the cognitive sciences*. Cambridge, MA: MIT Press.

Nakayama, K. (2005). Resolving border disputes in midlevel vision. *Neuron, 47,* 5–8.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception, 18*(1), 55–68.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Nguyen, J., Majmudar, U. V., Ravaliya, J. H., Papathomas, T. V., & Torres, E. B. (2016). Automatically characterizing sensory-motor patterns underlying reach-to-grasp movements on a physical depth inversion illusion. *Frontiers in Neuroscience, 9,* 694.

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 32/34

Nikara, T., Bishop, P. O., & Pettigrew, J. D. (1968). Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Experimental Brain Research, 6,* 353–372.

Ogle, K. (1950). Researches in *Binocular Vision*. Philadelphia, PA: Saunders.

Ogle, K. (1954). On stereoscopic depth perception. *Journal of Experimental Psychology, 48*(4), 225–233.

Ogle, K. (1959). The theory of stereoscopic vision. In S. Koch (ed.). *psychology: A study of a science, vol. I, sensory, perceptual and physiological formulations* (362–394). New York: McGraw Hill.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience, 8*(5), 379–391.

Parker, A. J. (2016). Vision in our three-dimensional world. *Philosophical Transactions of the Royal Society B, 371*(1697), 20150251.

Peacocke, C. (1983). *Sense and content: Experience, thought, and their relations.* Oxford: Oxford University Press.

Pentland, A. (1982). The perception of shape from shading. Proceedings of the OSA Annual Meeting, Oct. 18–22, Tuscon, AZ.

Pettigrew, J. D. (1965). *Binocular interaction on single units of the striate cortex of the cat*. Thesis, University of Sydney.

Poston, T., & Stewart, I. (1978). Nonlinear modeling of multistable perception. *Systems Research and Behavioural Science, 23*(4), 318–334.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*(3), 341–365.

Qiu, F. T., & von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. *Neuron, 47*(1), 155–166.

Ramachandran, V. S. (1986). Capture of stereopsis and apparent motion by illusory contours. *Perception & Psychophysics, 39*(5), 361–373.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331,* 163–166.

Ramachandran, V. S. (2006). Foreword. In A. Gilchrist (Ed.), *Seeing black and white*. Oxford: Oxford University Press.

Ramachandran, V. S., & Cavanagh, P. (1985). Subjective contours capture stereopsis. *Nature, 317,* 527–530.

Ramachandran, V. S., & Rogers-Ramachandran, D. (2009). Two eyes, two views: Your brain and depth perception. *Scientific American*, 1st Sept. 2009.

Rock, I., & Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American, 263*(6), 84–90.

Rogers, B., & Gyani, A. (2010). Binocular disparities, motion parallax, and geometric perspective in Patrick Hughes's 'reverspectives': Theoretical analysis and empirical findings. *Perception, 39*(3), 330–348.

Layout: **A5 HuSSci** Book ID: **455577_1_En** Book ISBN: **978-3-319-66293-0**
Chapter No.: **1** Date: **11 November 2017 14:31** Page: **33/34**

Rosas, P., & Wichmann, F. A. (2011). Cue combination: Beyond optimality. In J. Trommershäuser, K. Körding, & M. Landy (Eds.), *Sensory cue integration*. Oxford: Oxford University Press.

Sacks, O. (2006). Stereo Sue. In O. Sacks (Ed.), *The minds eye (2010)*. New York: Picador.

Sacks, O. (2010). Persistence of vision: A journal. In O. Sacks (Ed.), *The minds eye*. New York: Picador.

Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision, 11*(7), 1–17.

Schlosberg, H. (1941). Stereoscopic depth from single pictures. *The American Journal of Psychology, 54*(4), 601–605.

Seydell, A., Knill, D. C., & Trommershäuser, J. (2011). Priors and learning in cue integration. In J. Trommershäuser, K. Körding, & M. Landy (Eds.), *Sensory cue integration*. Oxford: Oxford University Press.

Sprague, W. W., Cooper, E. A., Reissier, S., Yellapragada, B., & Banks, M. S. (2016). The natural statistics of blur. *Journal of Vision, 16*(10), 23, 1–27.

Strawson, P. F. (1979). Perception and its objects. In G. F. Macdonald (ed.), *Perception and identity: Essays presented to A. J. Ayer, with his replies* (41–60). London: Macmillan.

Tausch, R. (1953). Die beidäugige Raumwahrnehmung. *Zeitschrift fur Experimentelle und Angewandte Psychologie, 3*, 394–421.

Todd, J. T., Egan, E. J., & Kallie, C. S. (2015). The darker-is-deeper heuristic for the perception of 3D shape from shading: Is it perceptually or ecologically valid? *Journal of Vision, 15*(5), 2.

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

Turner, R. S. (1994). *In the eye's mind: Vision and the Helmholtz-Hering controversy*. Princeton, NJ: Princeton University Press.

Tye, M. (1993). Blindsight, the absent qualia hypothesis, and the mystery of consciousness. *Royal Institute of Philosophy Supplement, 34*, 19–40.

Tyler, C. W. (1998). Diffuse illumination as a default assumption for shape-from-shading in the absence of shadows. *Journal of Imaging Science and Technology, 42*(4), 319–325.

Tyler, C. W. (2004). Theory of texture discrimination of based on higher-order perturbations in individual texturesamples. *Vision Research, 44*(18), 2179–2186.

Tyler, C. W. (2006). Spatial form as inherently three dimensional. In M. R. M. Jenkin & L. R. Harris (Eds.), *Seeing spatial form*. Oxford: Oxford University Press.

Vishwanath, D. (2005). The epistemological status of vision science and its implications for design. *Axiomathes, 15*(3), 399–486.

Vishwanath, D. (2010). Visual information in surface and depth perception: Reconciling pictures and reality. In L. Albertazzi, G. van Tonder, & D. Vishwanath (Eds.), *Perception beyond inference: The informational content of visual processes*. Cambridge, MA: MIT Press.

Layout: A5 HuSSci Book ID: 455577_1_En Book ISBN: 978-3-319-66293-0
Chapter No.: 1 Date: 11 November 2017 14:31 Page: 34/34

Vishwanath, D. (2016). Induction of monocular stereopsis by altering focus distance: A test of Ames's hypothesis. *i-Perception, 7*(2), 1–5.

Vishwanath, D., & Hibbard, P. B. (2013). Seeing in 3D with just one eye: Stereopsis without binocular vision. *Psychological Science, 24*(9), 1673–1685.

Volcic, R., Vishwanath, D., & Domini, F. (2014). Reaching into pictorial spaces. *Proceedings of SPIE, 9014: Human Vision and Electronic Imaging XIX*.

von der Heydt, R. (2015). Figure-ground organization and the emergence of proto-objects in the visual cortex. *Frontiers in Psychology., 6,* 1695.

von Rohr, M. (1903). The verant, a new instrument for viewing photographs from the correct standpoint. *The Photographic Journal, 43,* 279–290.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., et al. (2012a). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin, 138*(6), 1172–1217.

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P., et al. (2012b). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138*(6), 1218–1252.

Wagemans, J., van Doorn, A. J., & Koenderink, J. J. (2010). The shading cue in context. *i-Perception, 1,* 159–177.

Wallis, G., & Ringelhan, S. (2013). The dynamics of perceptual rivalry in bistable and tristable perception. *Journal of Vision, 13*(2): 24, 1–21.

Wertheimer, M. (1924). Ueber Gestalttheorie. Lecture before the Kant Gesellschaft.

Westheimer, G. (1994). The Ferrier Lecture, 1992. Seeing depth with two eyes: stereopsis. *Proceedings of the Royal Society B, 257*(1349), 205–214.

Westheimer, G. (2013). Clinical evaluation of stereopsis. *Vision Research*, 20 Sept. 2013, 38–42.

Wheatstone, C. (1838). Contributions to the theory of vision—Part the first, on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society, 128,* 371–394.

Wijntjes, M. W. A., Füzy, A., Verheij, M. E. S., Deetman, T., & Pont, S. C. (2016). The synoptic art experience. *Art & Perception, 4*(1–2), 73–105.

Wilde, K. (1950). Der Punktreiheneffekt und die Rolle der binokularen Querdisparation beim Tiefensehen. *Psychologische Forschung, 23,* 223–262.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in Monkey visual cortex. *Journal of Neuroscience, 20*(17), 6594–6611.