

Stereopsis in the Presence of Binocular Disparity

Abstract The strongest evidence that pictorial cues contribute to stereopsis is the fact that the visual system appears to integrate both pictorial cues (such as perspective and shading) and optical cues (such as binocular disparity) into a single coherent percept. Indeed, when these sources of information are slightly in conflict the visual system appears to construct an entirely new object that is not specified by any of the individual sources of information. But in this chapter I question whether what we experience in this context is really an *integrated percept*, as opposed to an *integrated judgment*, and I suggest experimental strategies that might enable us to distinguish between these two interpretations.

Keywords Cue conflict · Cue combination · Cue Integration
Mandatory fusion · Hollow-Face illusion · Reverspective

In Chap. 1, we saw how one of the strongest arguments against a purely optical account of stereopsis was the fact that pictorial cues appear to be able to modify the depth specified by binocular disparity. This effect was well documented in the mid-twentieth century by Ames (1951, 1955)

The original version of this chapter was revised: Post-publication corrections have been incorporated. The erratum to this chapter is available at https://doi.org/10.1007/978-3-319-66293-0_5

and Ogle (1959), and in the 1970s–1980s by Gregory’s (1970) hollow-face illusion and Hughes’ Reverspectives. But in the mid-1990s the ability of the visual system to reconcile conflicting sources of information into a single coherent percept became the *organising principle* of perception (see Landy et al. 1995; Knill and Richards 1996). You might reasonably wonder why? After all, the human visual system evolved in response to a cue-consistent real world rather than the artificially induced cue conflicts of the laboratory. But as Hillis et al. (2002) explain, contemporary articulations of Cue Integration start from the premise that every depth cue is subject to two sources of potential error, namely *bias* (inaccuracy) and *random noise* (imprecision), and this explains why estimates of the same property from different cues are liable to differ.

Furthermore, whilst some authors have explored potential *bias* (see Domini and Caudek 2011; Scarfe and Hibbard 2011), the majority of the Cue Integration literature proceeds on the basis that *random noise* is the most egregious concern. Hillis et al. (2002) are therefore typical when they assume that the visual system is well calibrated, so that signals will *on average* agree with one another. Instead, they assume that the source of any discrepancy is the random error that all measurements are subject to and which can be modelled by a Gaussian distribution. And it is upon this basis that Cue Integration has generally embraced a Bayesian weighted average as the most appropriate model of cue combination.

This Bayesian model of Cue Integration is typically evaluated by introducing conflicts between various depth cues and seeing if the visual system responds as expected. But one might reasonably question whether the cue conflicts employed in these experimental studies actually reflect a concern for random noise? For instance, it is hard to maintain that Ernst et al. (2000) (where 0° texture was pitted against 30° binocular disparity, and vice versa) or Hillis et al. (2002) (where $+20^\circ$ texture was pitted against -20° binocular disparity, and vice versa) merely modelled random noise in the visual system. Indeed, as Hillis et al. (2002) readily admit ‘such combinations rarely occur in the natural environment’; a point that Landy et al. (2011) reiterate: ‘one might argue that the artificial stimuli create cue conflicts that exceed those experienced under natural conditions...’.

Nonetheless, even in the context of these artificially accentuated cue conflicts, the visual system often appears to integrate depth cues in a linear fashion. So even if the Bayesian justification for these studies begins

to look questionable, their empirical findings, and especially the method used to procure them (the ‘perturbation analysis’ of inducing small cue conflicts: see Maloney and Landy 1989; Landy et al. 1991; Young et al. 1993), have become standard in the literature. Indeed criticism of Bayesian Cue Integration typically comes from those whose empirical findings in cue-conflict experiments are not consistent with a weighted average: see Domini and Caudek (2011) for an overview; and in particular Todd and Norman (2003), Likova and Tyler (2003), Vishwanath and Domini (2013), Vishwanath and Hibbard (2013), and Chen and Tyler (2015) for our present purposes.

But underlying this debate is the common assumption that the cue-conflict stimuli in these experiments really are integrated into a *single coherent percept*. This is true for those who advance a Bayesian account of Cue Integration, those who embrace an alternative conception of Cue Integration (see Domini and Caudek 2011; Tyler 2004), and even those who reject Cue Integration altogether (see Vishwanath 2005; Albertazzi et al. 2010; Koenderink 2010). For instance, although Vishwanath (2005) rejects Cue Integration, he nonetheless maintains that ‘cue-conflict stimuli are ideal for studying how co-calibration across sensory measurements is maintained: a calibration process that is designed to remove detected conflicts when possible.’ By contrast, the purpose of this chapter is to challenge the assumption that cue-conflicts are really eradicated at the level of *perception*. So whilst critics of Bayesian Cue Integration may challenge *how* these sources of information are perceptually integrated, I am asking the logically prior question of *if* they are perceptually integrated in the first place? But if they are not *perceptually* integrated, then what is the alternative? Well, the literature appears to draw a false dichotomy between (a) a *single integrated percept* and (b) *strategic decision-making*. For instance, in the context of vision and touch, Gepshtein et al. (2005) find evidence of subjects relying upon both sources of information, and ask: ‘Do the results manifest a unified multi-modal percept?’ And they admit that their results are silent between two competing interpretations:

The improvement in precision observed in the inter-modality experiment could in principle result from a perceptual process or a decision strategy.

And Gepshtein et al. clarify what each interpretation would entail

By the former, we mean that the observer's judgements are based on a unified multi-modal estimate resulting from the weighted combination of visual and haptic signals (Hillis, Ernst, Banks, & Landy, 2002).

By the latter, we mean that the observer's decision is based solely on comparing (and weighting appropriately) the two unimodal signals without actually combining them into a unified percept. That is, the information could still be used optimally, but without the percept of a single object.

Ultimately Gepshtein et al. conclude

Our study cannot distinguish between these two possibilities...

Similarly, when Todd and Norman (2003) observed that their subjects gave inconsistent evaluations of the stimuli depending upon how the stimuli were presented, they concluded that a strategic element must be at play:

The incompatibility of the objective data with the observers' phenomenal impressions provide strong evidence that there was a strategic component of their responses that was not based entirely on their conscious perceptions.

But I would argue that there is a third possibility, namely (as I outlined in Chap. 1) that rather than engaging in *conscious deliberation*, the subjects in Todd and Norman (2003) might simply be influenced by an *automatic* and *involuntary* cognitive process that operates *after* perception but *prior* to conscious deliberation. For instance, the attribution of meaning to words is not a *perceptual* process (it does not affect the visual appearance of the words themselves) and yet clearly operates *preconsciously* (we do not have to consciously attribute meaning to the words). But if a pre-conscious cognitive process can account for the attribution of meaning to words, why not the attribution of meaning to depth cues? Under this account the subjects' evaluations in Todd & Norman (2003) are not strategic but based upon an *integrated evaluative judgement* that had already occurred earlier in the cognitive chain. So the question we need to ask is whether cue-conflict stimuli really provide evidence for an *integrated percept* or merely an *unconscious post-perceptual integrated judgement*?

1 DOES CUE INTEGRATION CLAIM PERCEPTUAL FUSION?

One commentator has suggested that I have misinterpreted Cue Integration theory: they argue that Cue Integration is solely concerned with *performance*, rather than the basis of that performance, and so remains agnostic between these three different interpretations (perception, unconscious judgement, conscious decision strategy). I disagree for the following four reasons:

1. First, Cue Integration theorists are clearly cognisant of this argument. For instance, Held et al. (2012b) admit that in depth perception studies a conscious decision strategy based on a 2D interpretation of the cues is essentially always available, but generally unacknowledged. But what is interesting is that in their study of depth from defocus blur, Held et al. (2012b) reject this 2D interpretation by appealing to their subjects' visual experience: 'An important clue is subjects' phenomenology'. They asked their subjects whether they were relying on perceived depth or merely (as might be possible for defocus blur) a 2D inference and found that only one out of their four subjects relied on a 2D strategy, and even then only rarely. Consequently, they concluded that their findings must reflect *perception*.

Indeed, appeals to the subjects' own visual experience are made not only to confirm data that appear to be consistent with Cue Integration (such as Held et al. 2012b) but also, more controversially, to discount data that appear to contradict Cue Integration. For instance, when the subjects in Hillis et al. (2002) were able to discriminate stimuli that ought to be indiscriminable so far as Cue Integration is concerned: 'The participants' phenomenology was informative.' The subjects reported that the stimulus introduced 2D texture distortions that enabled them to discriminate between the stimuli, enabling Hillis et al. to maintain that the 3D cues to slant were truly fused, and that subjects only had access to a single depth percept, in spite of their contradictory performance.

2. Second, the very rationale of Cue Integration suggests that integration must *perceptual*: if the purpose of Cue Integration is to reduce the impact of random system noise by averaging across various noisy cues, why would the visual system give us direct access (via perception) to one of these noisy cues? The implication, as Hillis et al. (2002) explain, is that Cue Integration must not only have a *positive* dimension (improved performance when reliance on two or more cues would be beneficial),

but also a *negative* dimension (reduced performance when reliance on one cue alone would be beneficial). Hillis et al. (2002) term this negative dimension of Cue Integration *mandatory fusion*; specifically, a *loss of access* to individual depth cues: the visual system specifies a *single* depth estimate, which is beneficial from an evolutionary perspective (in a cue-consistent world, discrepancies are more likely to come from the visual system's inconsistent measurements), but which leads to detrimental performance in response to artificially contrived cue conflicts in the laboratory.

3. Third, *mandatory fusion* is an inevitable consequence of Cue Integration for another, more immediate, reason; namely, how Cue Integration conceives of the depth estimates from individual cues: It doesn't treat the sensory data as a *specific* estimate of depth, but rather as the basis for a probability distribution (a 'likelihood function') which plots the probability of receiving *this* sensory data from a variety of different potential depth values whose signal has been corrupted by noise. Consequently, there is no possibility that perception reflects *the* estimate from one specific cue, since there is no one specific estimate, only a set of probabilities.

4. Fourth, the final reason that we know Cue Integration is committed to *perceptual* integration is that many of its most startling claims are articulated as claims about *visual experience* rather than *performance*. Consider, for instance, Ernst et al.'s (2000) paper: 'Touch Can Change Visual Slant Perception'. Ernst et al. went beyond merely an observation about *performance*, namely that touch feedback can affect the weight given to various sources of visual information, to an observation about *visual experience*, namely that touch can change the slant that is *seen*. Indeed, as the title of their paper illustrates, it was this claim about *visual experience*, rather than the improved cross-modal *performance*, that proved to be the central message of their paper.

The claim that touch can influence the slant that is *seen* has been thrown into doubt by the subsequent literature. For instance, as the title of their paper suggests, Hillis et al. (2002) found 'mandatory fusion within, but not between, senses'. Nonetheless, even if we stick to Cue Integration within vision itself, mandatory fusion has quite profound implications for our visual experience: as Hillis et al. explain, an appropriately calibrated high-texture-low-disparity

stimulus and low-texture-high-disparity stimulus should be *perceptually indistinguishable*.

Indeed, Hillis et al. develop this point with an analogy from the colour literature, namely *metamers*: ‘composite stimuli that cannot be discriminated even though their constituents can be’. So just as red and green light added together is subjectively indistinguishable from yellow light, cue-conflict stimuli can be subjectively indistinguishable even though had their disparity or texture been presented in isolation you would be able to differentiate them. Indeed, Hillis et al.’s *metamer* analysis had such a profound effect on the literature that within two years, it was legitimate for Ernst and Bühlhoff (2004) to simply assume mandatory fusion as an initial premise rather than a conclusion that had to be argued for.

2 DOES CUE INTEGRATION DEMONSTRATE PERCEPTUAL FUSION?

But how do Hillis et al. (2002) justify their claim that Cue Integration produces *metamers*? I.e. that an appropriately calibrated high-texture-low-disparity stimulus and a low-texture-high-disparity stimulus are *perceptually* indistinguishable?

First, Hillis et al. take a cue-consistent stimulus (with the same slant specified by texture and disparity) and introduce a cue conflict by varying the stimulus along one of two dimensions until the subject is able to identify the altered stimulus (by picking the odd-one-out when the altered stimulus and two unaltered stimuli are shown in succession). The results of this preliminary study were then used to mark out each subject’s subjective thresholds for changes in texture and disparity (the parallel lines in Fig. 1, left), and the question was whether altering the stimulus along *both* dimensions at the same time could *improve performance* (with subjects noticing two complementary sub-threshold changes in disparity and texture: the *positive* dimension of Cue Integration), or even *worsen performance* (with subjects failing to notice an above-threshold change in the one cue if a change in the opposite direction is made in the other: the *negative* dimension of Cue Integration), as predicted by Hillis et al.’s (2002) model of mandatory fusion?

Figure 2 provides an illustration of the subjects’ performance in Hillis et al. (2002). It is unclear how representative these results are, but they suffice for the purposes of our discussion. They do tend to show the

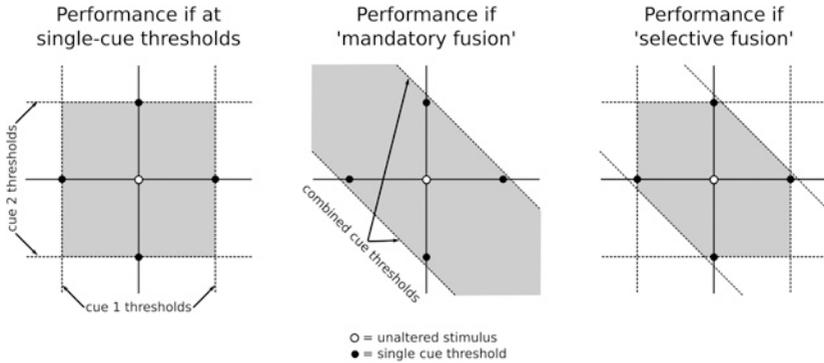


Fig. 1 Hillis et al.'s (2002) predictions if subjects **a** use single-cue estimates (no Cue Integration), **b** only have access to a combined estimate (mandatory fusion), or **c** experience the benefits of Cue Integration without the costs of mandatory fusion (selective fusion)

predicted *improvement* in performance, but evidence for the predicted *deficit* in performance seems patchy: certainly subject JH seems to be performing close to threshold, as does subject AH on occasion.

Hillis et al. take those instances where there was a performance deficit as evidence of mandatory fusion. But they recognise that mandatory fusion should not be partial, and so try to explain why the predicted deficit wasn't always present. As mentioned above, they suggest that the texture of the stimulus was subject to distortions as disparity was increased, and it was on this basis (rather than 3D slant) that subjects were able to identify the odd-one-out. But there are two concerns with this explanation: The first is that this psychophysical task was chosen as one that would establish mandatory fusion via performance without the need to evaluate experience, so it is concerning to see Hillis et al. using subjective experience to explain away performance that is contrary to their hypothesis. The second is that it is unclear to what extent their explanation maps their results: First, why did it not affect those trials where the predicted performance deficit was found? Second, why does performance from texture distortion (i) coincide almost exactly with the subject's own single-cue thresholds, and (ii) not appear to change significantly as disparity is increased?

Still, we have to explain those instances where the performance deficit predicted by mandatory fusion was present. Is this the clear evidence of mandatory fusion that Hillis et al. suggest?

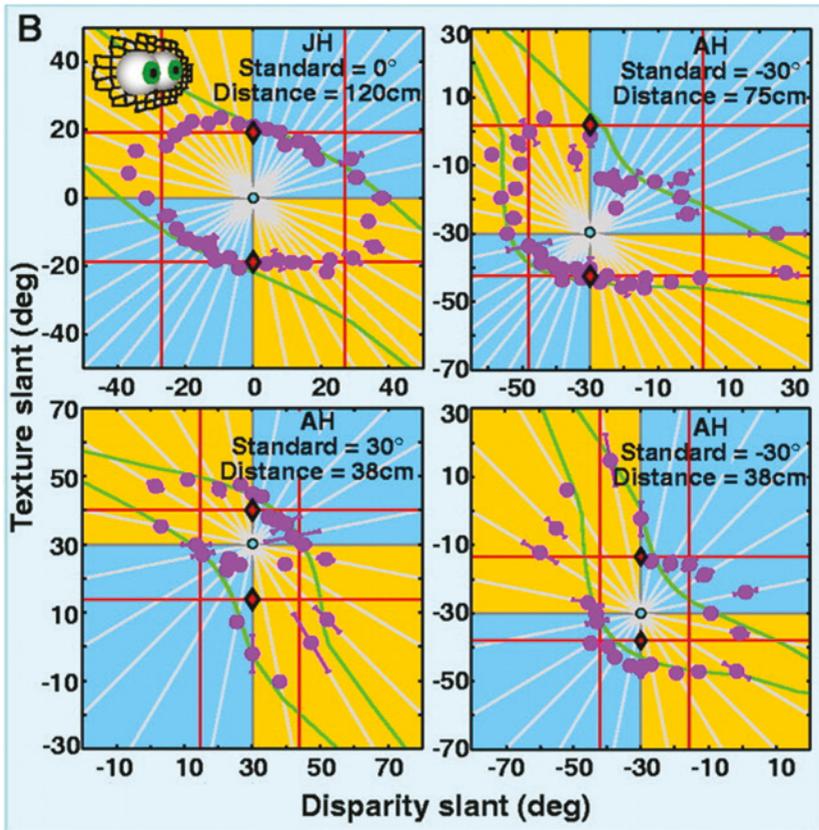


Fig. 2 Illustrative results from two participants in Hillis et al. (2002): The two sets of parallel lines represent the subject's single-cue thresholds. Any point within the rectangle demarcated by the parallel lines represents an improvement over single-cue performance, whilst any point outside the rectangle demarcated by the parallel lines represents a performance deficit relative to single-cue performance. The curves represent Hillis et al.'s model of the optimal combined estimator (Fig. 1, middle) taking into account how the weights assigned vary with texture-specified slant. From Hillis et al. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, 298, 1627–1630. © The American Association for the Advancement of Science

1. Direct Comparison: My primary concern with Hillis et al. isn't the partial nature of their results, but what their results are evidence for. The problem is that in evaluating perception Hillis et al. introduce a memory component: their three stimuli are presented sequentially for 1.5 s with a 0.3 s interval between them. So rather than claiming that their data provide a clear demonstration of single fused *percept*, I would suggest that their results are only evidence of a single fused *memory*.

That our *memory* of the slant of a percept is based on a single overall estimate or impression is entirely plausible. As Cavanagh (2011) observes:

Clearly, the description of visual scene cannot be sent in its entirety, like a picture or a movie, to other centers as that would require that each of them have their own visual system to decode the description. Some very compressed, annotated, or labelled version must be constructed that can be passed on in a format and that other centers – memory, language, planning – can understand.

But equally, as this quotation illustrates, we cannot simply presume that just because *memory* operates in this way, that so too must *perception*.

But what is the alternative? Well, it might be that sequential tasks (asking subjects to make a comparison between stimuli *over time*) are simply an inappropriate basis upon which to evaluate *perception* rather than *memory*. And we should not shy away from this conclusion if that is what the logic of the distinction between *perception* and *memory* requires. That being said, I do think that we can legitimately question why Hillis et al. introduce a 0.3 s interval between their stimuli? If the various stimuli really are *metamers* in the strong sense that Hillis et al. suggest, and is implied by mandatory fusion, then we have to question why this 0.3 s interval is required: for instance, if we wanted to demonstrate that two shades of yellow were subjectively *indistinguishable*, we would simply alternate between them without an interval, so why should 3D depth be any different?

One response, suggested to me by a commentator, is that *motion* might be processed separately from *3D form*: so whilst *3D form* might be indistinguishable, motion detectors may alert the subject that *something* has changed in the stimulus, even though the subject cannot identify what this change was. In one sense, this problem is caused by Hillis et al.'s reliance on *performance* to determine mandatory fusion, leaving

subjects free to rely on any means possible to pick the odd-one-out. But do concerns about motion detection from removing the inter-stimulus interval concede too much?

First, the whole point of Cue Integration is that we give *meaning* to noisy cues. If it would be unwise for the visual system to give subjects access to individual noisy cues in the context of *3D form* perception (the mandatory fusion thesis), then why would it make any more sense in the context of *motion*? After all, according to the perturbation analysis, we are meant to be modelling random noise in the visual system. And if motion detectors were triggered every time random noise in the visual system fluctuated, this would be a recipe for evolutionary disaster.

Second, even if subjects could tell the difference between the two stimuli with the interval removed, it could still be a good test so long as we reintroduce an *evaluative* component and asked subjects whether the impression of depth between the two stimuli was *qualitatively* similar? Even if subjects judge the two stimuli to have the same *quantity* of depth, do they really lose nothing (so far as perceptual depth is concerned) as we alternate between them?

In conclusion, we want to avoid reducing our evaluation of visual depth into a change-blindness paradigm, so the choice seems clear: either we test the sequential paradigm *without* an artificially induced interval, or we avoid the sequential paradigm altogether.

2. Subjective Evaluation: At the same time, Hillis et al. (2002) were clearly onto something when they sought to eradicate an *evaluative* element from their task. Asking people whether a stimulus *looks flat*, *looks slanted*, or *looks bulged*, is as much an evaluative judgement as asking someone if a stimulus *looks square* or *looks symmetrical*. So how do we know that pictorial cues contribute to our perception, rather than merely biasing our evaluation?

Indeed, under my account (where the 3D form of a cue-conflict stimulus is specified solely by its binocular disparity), there are good evolutionary reasons for divorcing our *evaluation* of the scene from our *perception* of it: binocular disparity reduces with distance, but the physical geometry of the scene does not. Consequently, if we wish to use our evaluations as the basis for our interactions with the invariant physical world, we cannot rely too heavily upon our perceptual impression of stereopsis. Indeed, this concern continues to apply (albeit with less force) in the context of linear Cue Integration, where the reduction of binocular disparity with distance still affects the perceived depth of the scene. Nor should we be

surprised that our *perception* and *evaluation* of the depth in a scene can come apart: we are quite capable of watching TV at close quarters (e.g. on a laptop screen: 40–50 cm, or even a phone: 30 cm) without the flatness specified by binocular disparity significantly impeding our enjoyment, and this might explain the indifference that the general public has recently shown towards 3D movies.

Like Hillis et al. (2002), the purpose of this chapter is to try and identify ‘a true test for the existence of cue fusion.’ And like Hillis et al., I am sceptical that relying on subject’s evaluative judgements provides that evidence. To see why, consider Ernst et al. (2000). The subjects were shown cue-conflict stimuli with inconsistent slants specified by texture and disparity (Fig. 3). The subjects received touch feedback that was consistent with texture or disparity, and this touch feedback influenced the estimate of slant. But none of the subjects in the experiment noticed that either (a) the slants specified by texture or disparity were different, or (b) that the touch feedback was consistent with one but not the other. So if we were to determine mandatory fusion simply by asking subjects for their subjective impressions we would have at least one *false positive* in this case: as Hillis et al. (2002) have convincingly demonstrated, there is *no* mandatory fusion in the cross-modal context of vision and touch. So the cross-modal integration in Ernst et al. (2000) must simply reflect the subjects’ *post-perceptual evaluation* of the stimulus. But in which case, what makes us any more confident that the integration of texture and disparity in the unimodal case of vision is any more *perceptual*? As we learnt from the cross-modal context, the fact that they might *seem* integrated is not enough.

More evidence that we cannot simply delegate this question to subjects’ own subjective evaluations comes from Todd and Norman (2003). Todd and Norman asked their subjects to evaluate the depth from (a) a monocular motion display, (b) a static binocular disparity display, and (c) a binocular disparity plus motion display, and found that depth was judged to be highest in the monocular motion display and lowest in the static binocular disparity display, with the binocular disparity plus motion display falling midway between the two. Indeed, all subjects judged the binocular disparity plus motion display to have at least 15% less depth than the monocular motion display. But Todd and Norman asked their subjects to close one eye as they watched the binocular disparity plus motion display, and report whether they saw an increase or a decrease in depth? All the observers reported a significant reduction in

the perceived depth, even though closing one eye converts the binocular disparity plus motion display into the monocular motion display that had earlier been evaluated as having 15% more depth.

Todd and Norman correctly conclude that, out of the two results, the direct and immediate comparison of closing one eye gives us a truer impression of actual perception than the subjects' own evaluations. In short, if subjective evaluations are liable to reverse the depth order of stimuli from 'monocular < binocular' to 'monocular > binocular' then we have good reason to be sceptical of them.

But we still have to explain why the subjective evaluations of depth reversed the depth order? As we have already discussed, Todd and Norman suggest that the subjective evaluations had a strategic component. Specifically, they claim that their subjects had to consciously convert their perceived depth into physical depth by comparing it to the height and width of the displays. But the problem with this explanation is that this concern equally applies to both the monocular motion and the binocular disparity plus motion displays, so it doesn't explain why the translation of perceived depth into physical depth should have reversed the depth order between the monocular motion and binocular disparity plus motion displays.

Instead, I would argue that the subjects simply *misjudged* their own visual experience in the monocular motion display: they thought they saw more depth than they actually did. This is because their evaluation of own their perceptual experience is, like any cognitive process, open to being *biased* or *prejudiced* by the depth *depicted* by the monocular cues. As I explain in Chap. 3, we can only know how much depth a non-disparity display produces by viewing it synoptically (sending an identical image to both eyes) and then introducing various points with binocular disparity into the scene. And this is essentially what Todd and Norman got their subjects to do in reverse by closing one eye, with the depth from disparity throwing the comparative flatness of the pictorial cues into sharp contrast.

Indeed, Todd and Norman (2003) provide us with a valuable illustration of the dilemma facing the Cue Integration literature: either we (a) rely on subjective evaluations, in which case there is no guarantee that the subjects' evaluations of their own perceptual experience is accurate (and, following Ernst et al. 2000; Todd and Norman 2003, significant evidence that it is not), or (b) we attempt to make a direct comparison between the stimuli, which may work in some contexts (e.g.

Todd and Norman 2003), but which may raise apparent motion concerns others (e.g. Hillis et al. 2002).

3. Indirect Comparison: But what we are studying is not merely the mechanisms that underpin depth perception in the laboratory, but also the mechanisms that explain our perception of the real world. And in the real world, we rarely get a chance to view and evaluate objects in isolation; we have no choice but to gauge their geometry in the presence of other objects. So although there is evidence that *proximity* (Gogel 1956) and *framing* (Eby and Braunstein 1995) may influence our evaluations when objects are not viewed in isolation, these influences cannot be so pervasive that they render any such evaluation completely redundant. Which opens up the possibility of a *third* strategy:

Instead of asking subjects to (a) *directly compare* Stimulus A with Stimulus B, or (b) *subjectively evaluate* Stimulus A in isolation, and then Stimulus B in isolation, and then compare these evaluations, we might attempt to (c) *indirectly compare* Stimulus A to Stimulus B, by first comparing Stimulus A to Stimulus C and then Stimulus B to Stimulus C. Indeed, Stimulus C might well be a second object or visual element that persists *at the same time* as both Stimulus A and Stimulus B. Nor should *proximity* or *framing* overly concern us; to the extent these concerns are brought into play they ought to equally affect the comparison between Stimulus A and Stimulus C on the one hand and Stimulus B and Stimulus C on the other.

But what would be a suitable comparator? Well, given binocular disparity is a cue to depth *off the fronto-parallel plane*, it would be useful to have an object or cue that marked out the location of the fronto-parallel plane, against which we could judge the degree of stereopsis in the scene with a simple comparison. Ironically enough, just such a cue was introduced by Ernst et al. (2000) in Fig. 3. Notice the black crosses in the centre of the stimuli: these black crosses were not present in the original experiment, but were added to the published version of the stimuli to help readers cross-fuse. But in the context of our discussion these black crosses take on another role: since they lack disparity or perspective, they demarcate the fronto-parallel plane. Nor should the presence of these black crosses overly affect the cue-conflict stimuli themselves: both the crosses and the cue-conflict stimuli ought to be regarded as freestanding objects defined by their own perspective and disparity cues. Nor should the crosses affect one cue-conflict stimulus more than the other: if Cue Integration really does occur *prior* to form perception, it shouldn't

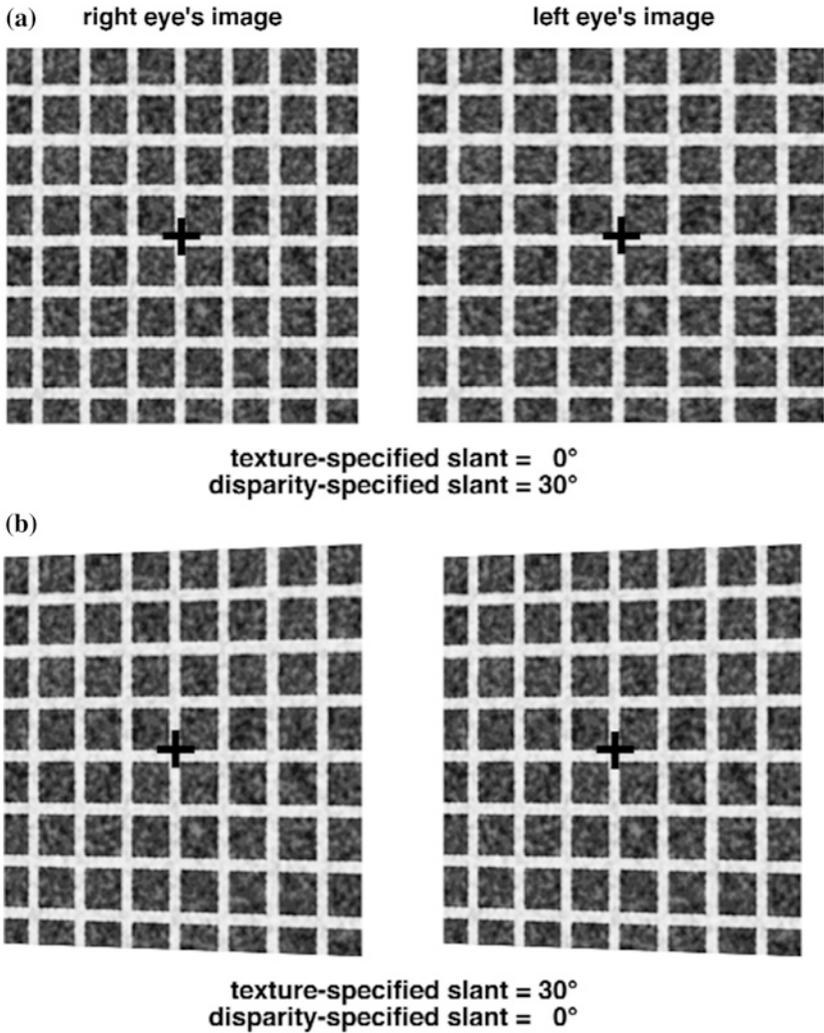


Fig. 3 Two examples of the cue-conflict stimuli from Ernst et al. (2000): **a** Texture specifies a slant of 0° whilst disparity specifies a slant of about 30° (at a viewing distance of 20 cm). **b** Texture specifies a slant of about 30° whilst disparity specifies a slant of 0° . From Ernst et al. (2000). Touch can change visual slant perception. *Nature Neuroscience*, 3(1), 69–73. © Nature Publishing Group

matter whether the perceived slant is primarily a product perspective or disparity; the only thing that matters is the overall all-things-considered determination.

And yet, even with all these provisos in place, this isn't how we experience the stimuli: in Fig. 3a, the cross is clearly slanted in stereoscopic space against the stimulus, whilst in Fig. 3b the cross is clearly flat against the stimulus. Indeed, this observation is only accentuated when we elongate the horizontal bars of the cross in (Fig. 4). So according to the quick and easy comparison that the black crosses and the horizontal bars afford us, the binocular disparity in Fig. 3a contributes positive stereoscopic slant, but the perspective cue in Fig. 3b does not. Nor do I think that this is an artefact of introducing the black crosses or horizontal bars: admittedly the fact that we have to read the stimulus as transparent in order to reconcile the slant of the cross with the slant of the stimulus might introduce some complexity, but this interpretation is readily adopted in Fig. 3a, so why not Fig. 3b?

As we have already observed, if Cue Integration is truly a *perceptual* phenomenon then we would expect it to be robust enough to survive interaction with other objects. But if, as appears to be the case, the depth specified by Cue Integration evaporates as soon as it comes into contact with another object, we have to seriously question whether it was really there in the first place. Ernst et al. (2000) suggest that Cue Integration occurs in Fig. 3b because 'most viewers perceive a slant between 0° and 30° , because both signals contribute to the perceived slant'. But we could just as easily imagine the perspective cues in Fig. 3b biasing the subjects' *evaluation* of the depth they perceive, but not their actual *perception*.

Finally, Fig. 3a illustrates another concern for Ernst et al.'s (2000) method of *subjective evaluation*: Glennerster et al. (2002), Glennerster and McKee (2004) observe that when subjects infer a frame of reference for their stereoscopic judgements it is rarely the fronto-parallel plane. This is demonstrated by Fig. 3a, where it is the cross that looks slanted relative to the stimulus, and not the other way round, further demonstrating just how poor our ability to evaluate stereoscopic depth really is.

To return to the question of pictorial cues biasing our *subjective evaluation* of stereoscopic depth, I would argue that the very same effect is evident not only when subjects (a) attribute depth to pictorial cues in the absence of binocular disparity (as in Ernst et al. 2000), but also (b) when subjects fail to attribute depth to small but otherwise discriminable

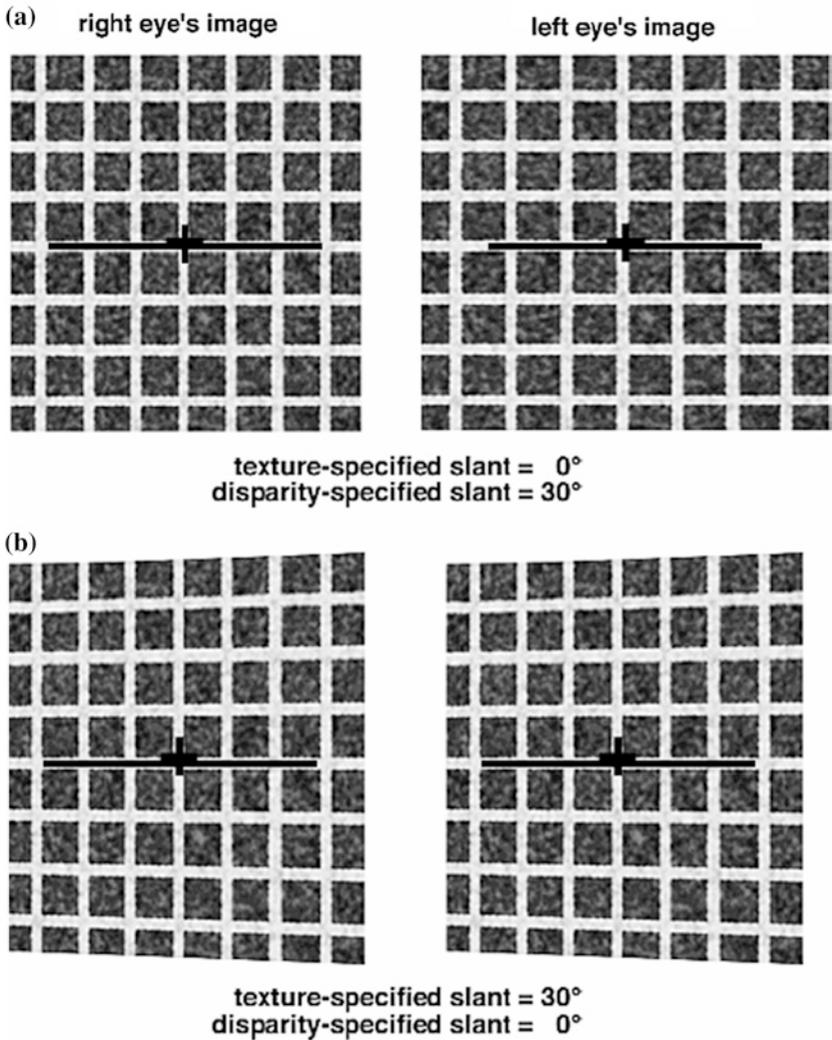
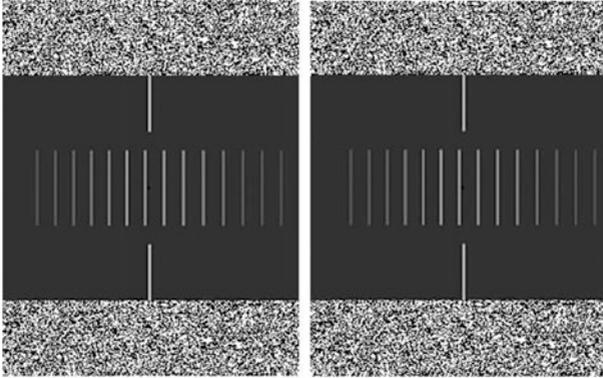
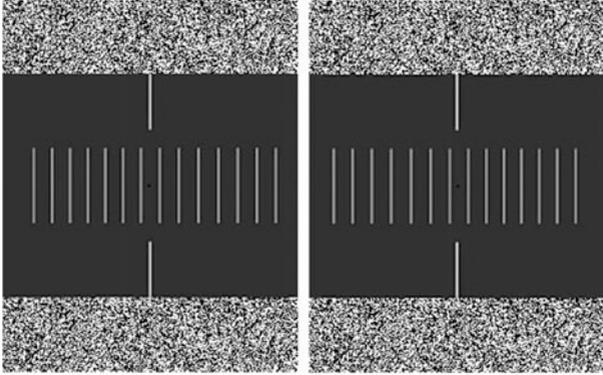


Fig. 4 Fig. 3 with a horizontal bar added. Amended from Ernst et al. (2000). Touch can change visual slant perception. *Nature Neuroscience*, 3(1), 69–73. © Nature Publishing Group

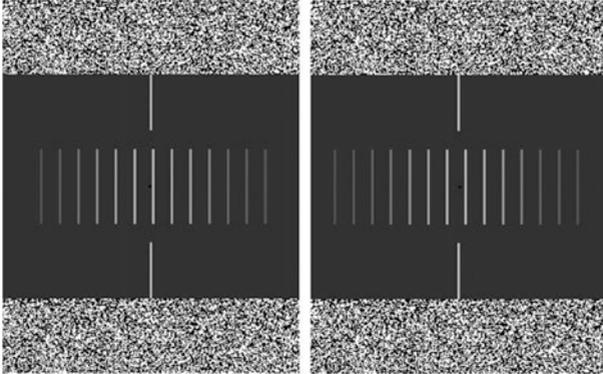
(a)



(b)



(c)



- ◆ **Fig. 5** Stimuli from Likova and Tyler (2003): sparsely sampled Gaussian profiles defined by **a** luminance only, **b** disparity only, and **c** ‘a combination of both cues at the level that produced a cancellation to flat plane under the experimental conditions’. From Likova and Tyler (2003). Peak localization of sparsely sampled luminance patterns is based on interpolated 3D object representations. *Vision Research*, 43, 2649–2657. © Elsevier

binocular disparities when conflicting pictorial cues are present, such as in Likova and Tyler (2003). Now there are two distinct questions in Likova and Tyler (2003) that I want to keep separate:

The first is how good we are at filling in on the basis of sparse information? For instance, if I give you the following sequence: 1, 2, 3, ..., 5, ..., 7, ..., you might immediately see that the missing numbers are 4, 6, and 8. And if I asked you to pick the highest value in the sequence you would point to the last ..., even though its value is not explicitly specified. Now according to Likova and Tyler (2003), it turns out that we are much better at inferring *3D form* from sparse information (e.g. Gregory’s dalmatian: see Fig. 4 in Chap. 3) than we are at inferring *changes in surface luminance* from sparse information (e.g. Hume 1748’s ‘missing shade of blue’ in a sequence of blue patches of increasing luminance); and Likova and Tyler demonstrate this fact with Fig. 5:

Consider the 14 vertical bars with varying luminance in the two (identical) images in Fig. 5a. We could interpret them as 14 individual vertical bars, or we could interpret them as part of a single continuous horizontal black-and-white surface. Even as a horizontal surface, Fig. 5a is open to two interpretations: a flat 2D surface with a change in luminance or a convex 3D surface cast in shadow.

First, Likova and Tyler found that in spite of the absence of foreshortening (see Hartle and Wilcox 2016), subjects automatically adopted the latter (3D) interpretation: the luminance profile evoked ‘an unambiguous depth percept of the brighter bars appearing closer for all observers’. Second, Likova and Tyler found that subjects were actually quite good at interpolating the location of the surface bulge when it was interpreted as a bulge in 3D shape. Third, however, when this 3D interpretation was barred by a competing disparity profile, and subjects were left trying to interpolate the bulge as a change in the luminance of a 2D surface, they were unable to perform the task: ‘Once the depth interpretation is nulled

by the disparity signal, the luminance information does not support position discrimination at all’.

Now the second question that Likova and Tyler explore, and the one that concerns us, is their use of disparity to cancel or null the 3D interpretation of the luminance profile, leaving a surface with otherwise discriminable luminance and disparity cues looking flat. Likova and Tyler confirm this cancellation effect by testing the disparity at which the luminance profile looked flat, and found that this did not occur at zero disparity, but at a small negative disparity of between -0.3 and -0.4 arc min. For Likova and Tyler, this observation confirms the fact that ‘the perceived depth from the luminance profile lies in the same qualitative dimension as the perceived depth from disparity cues (i.e. that it is a ‘true’ depth percept rather than just a cognitive inference of some kind)’. Now if by a ‘cognitive inference’ Likova and Tyler mean that subjects consciously subtract the depth from disparity from the depth from luminance (for instance, Likova and Tyler allude to the possibility that ‘the luminance patterns might be interpreted as an object during localization’), then I quite agree. But as I have continually emphasised in this chapter, between the *perceptual bias* that Likova and Tyler argue for, and the *conscious decision-making* that Likova and Tyler reject, there is a third possibility, namely a *cognitive bias*: the idea that our evaluation of our own visual experience can be biased by the presence of confounding cues.

Such an interpretation is entirely consistent with Likova and Tyler’s cancellation paradigm, especially since (a) the disparity that is cancelled is small (0.3 – 0.4 arc min), and (b) the determination of the null-point itself relies upon a cognitive judgement of comparative depth: as Likova and Tyler explain, the null-point is only approximate, and required subjects to judge whether the centre of the bulge appeared to be at the same depth as the bars on the far left and far right, even if some ‘minor wrinkles’ could be seen in the transition regions. But as I constantly emphasise, the judgement that something ‘looks flat’ or ‘looks bulged’ is exactly that: a *judgement*; and, as with all judgements, we need to ask what confidence we have in our ability to make these judgements without bias?

My interpretation would also be consistent with the results of Likova and Tyler’s main study where the luminance profile introduced a small but consistent bias in favour of concave depth, but left depth from disparity otherwise intact (Fig. 6): As the curves fitted through the points in Fig. 6 (a) and (b) demonstrate, the effect of luminance is to shift the

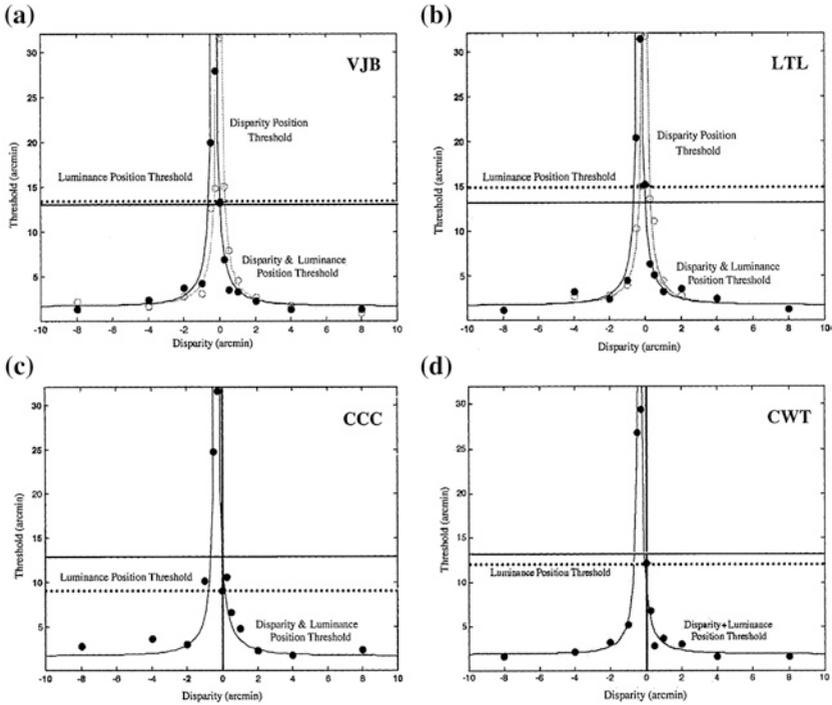


Fig. 6 The results of the position localisation task for the two principal observers in Likova and Tyler (2003): VJB and LTL, with key conditions verified with another two observers: CCC and CWT. The white circles are the thresholds for the disparity only condition, and the black circles are the thresholds for the disparity plus luminance condition. From Likova and Tyler (2003). Peak localization of sparsely sampled luminance patterns is based on interpolated 3D object representations. Vision Research, 43, 2649–2657. © Elsevier

whole psychometric function to the left by 0.3–0.4 arc min. As Likova and Tyler observe, this is the only change that luminance makes: ‘all other aspects of the position task fell on the same curve with no change in parameter values’. So the results clearly demonstrate a fixed *bias* in extracting depth from disparity. But they don’t determine whether this bias is *perceptual* or *cognitive*.

Third, there is one basis upon which we might try to determine whether the bias in Likova and Tyler is *perceptual* or *cognitive*. According to Likova and Tyler, the long-range interpolation process that forms

the basis of their study is the *only* means by which we can see everyday objects: they argue that since everyday objects are typically defined by local features separated by extended featureless regions the visual system has to engage in long-range interpolation to extract their 3D form. But if the integration of disparity and shading is really how we see everyday objects, then we would expect Likova and Tyler's stimuli to behave like ordinary visual objects: bringing an independent object into the vicinity of the surfaces shouldn't turn a *convex* surface *flat* and a *flat* surface *concave*.

And yet this is exactly what appears to happen when we place their stimuli alongside a reference point or alongside one another. In their actual experiment, Likova and Tyler offset the disparity of the reference point as a way of ensuring that subjects were interpolating the left-right location of the bulge. By contrast, the question that concerns us is whether luminance nulls depth from disparity? I.e. whether there is a 3D bulge in the first place. And in this context, there is no harm in setting the reference point at zero disparity as a test of true flatness: rather than engaging in what is, by Likova and Tyler's own admission, a long-range evaluative judgement in order to judge the presence or absence of flatness (by comparing the depth of the central bar against the bars on the far left and far right), what harm could it do to afford subjects a closer reference point in order to make their determinations?

But, as soon as we do, the *concavity* specified in Fig. 5c by binocular disparity (0.4 arc min) becomes fully apparent (Fig. 7).

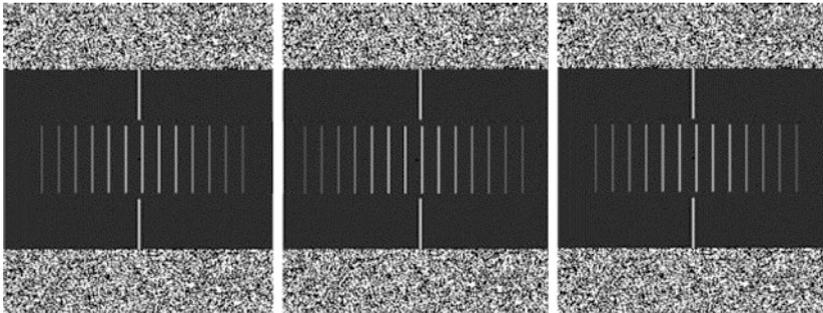


Fig. 7 Fig. 5c with disparity removed from the reference point, and the reference point brought closer to the stimulus. Amended from Likova and Tyler (2003). Peak localization of sparsely sampled luminance patterns is based on interpolated 3D object representations. Vision Research, 43, 2649–2657. © Elsevier

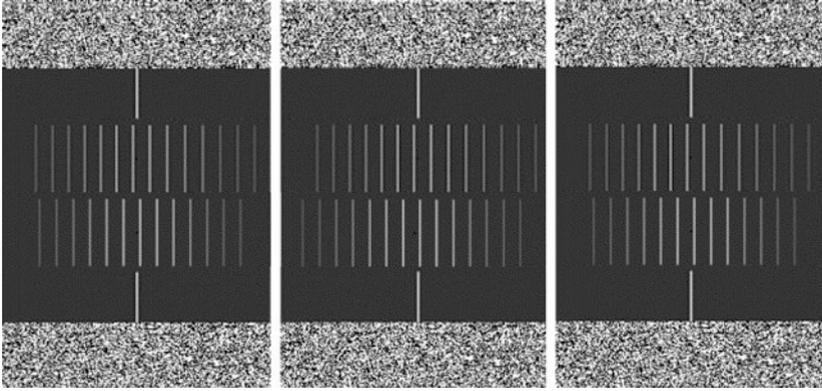


Fig. 8 Fig. 5a (top) added to Fig. 5c (bottom) with disparity removed from the reference point, and the reference and the stimuli brought closer together. Amended from Likova and Tyler (2003). Peak localization of sparsely sampled luminance patterns is based on interpolated 3D object representations. Vision Research, 43, 2649–2657. © Elsevier

Furthermore, the *flatness* specified in Fig. 5a by the absence of binocular disparity also becomes immediately apparent when it is added (upper stimulus) (Fig. 8).

None of the stimuli are occluded, so each ought to persist in its own depth defined by its own depth cues. So why, once we afford subjects a more accurate reference point by which to judge the *presence* or *absence* of stereoscopic depth, does depth from shading appear to evaporate?

3 ILLUSIONS

Having altered the cue-conflict stimuli in Ernst et al. (2000) and Likova and Tyler (2003) in order to better understand their *perceived* rather than merely *conceived* depth, we might wonder how a similar technique would affect the real-world cue conflicts encountered in Reverspectives (Fig. 4 in Chap. 1) and the hollow-face illusion? For instance, if we add

horizontal and/or vertical bars to these illusions, what happens? Do the bars cut through the illusory depth? Or do they break the illusion altogether? (Figs. 9, 10)

In fact, neither occurs. Instead, the illusion persists *but the inverted depth is located behind the horizontal and vertical bars*: the inverted depth does not protrude beyond the bars even though as an inverted depth percept it ought to. It is as if the *inverted percept* and *stereoscopic space* are simply talking past one another. And I would argue that the only way to make sense of disconnect is to recognise that the inverted depth percept is, in fact, a *false judgement* that we apply to a *veridical percept* of the hollow face or Reverspective. In which case, the hollow Face and Reverspectives are better thought of as *delusions* rather than *illusions*: *misinterpretations* of what we see, rather than *false percepts*.

This observation opens up a whole new experimental strategy in trying to understand Reverspectives and hollow-face illusion. Indeed, we can place objects not just *in front*, but also at various points *inside*, the Reverspective and the hollow face, in space that ought not to exist according to the illusory percept. Does this destroy the illusion? Again, it doesn't appear to: swaying back and forth, we still get the 'illusory percept', in spite of the fact that we are also aware that we are viewing a hollow filled with objects. Indeed, we might distribute points within the hollow space, or place an object with an identifiable slant, in order to see how, if at all, our ordinal depth judgements are affected? For instance, if we place a pen at a slant in the hollow of the Reverspective, my experience is that we continue to see it as slanted in the right direction even though the 'illusory percept' persists (Fig. 11).

This effect would have to be confirmed experimentally, and perhaps the best test would be a giant (1.5–2 m) Reverspective or hollow face: on the one hand, the binocular disparity of the structure will be reduced, so we should be able to get closer whilst still maintaining the *global* inverted depth illusion, but on the other hand, if we use markers distributed in space the depth separation between these points will be increased, thereby accentuating the *ordinal* depth we have to judge.

Admittedly, we might see the pen move in Fig. 11, so do we at least get an illusory percept of motion, if not an illusory percept of depth? I would resist this conclusion not only because illusory depth and illusory motion appear to be two sides of the same coin (if one is a judgement,

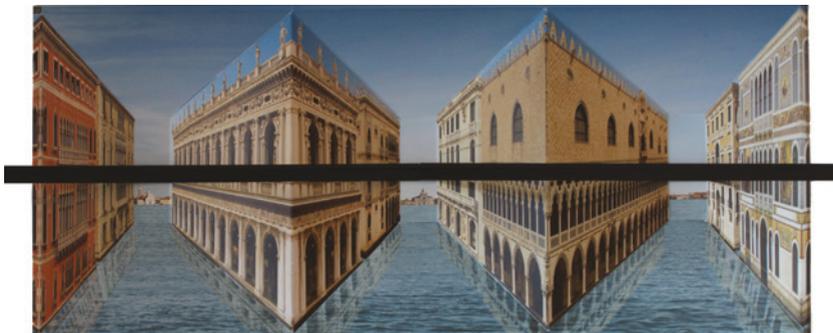


Fig. 9 Reverspective with horizontal bar attached. Reverspective courtesy of <http://www.offthewallartprints.com>. © Off The Wall Art Prints



Fig. 10 Hollow-face illusion with vertical and horizontal bars attached. Hollow-face illusion courtesy of <http://www.grand-illusions.com>. © Grand Illusions

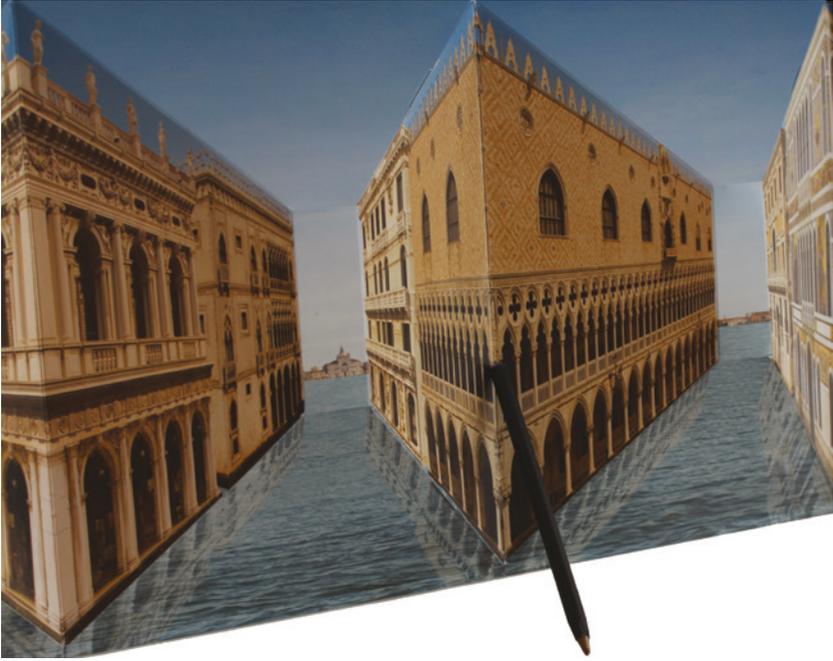


Fig. 11 Pen placed in the recess of a Reverspective. Reverspective courtesy of <http://www.offthewallartprints.com>. © Off The Wall Art Prints

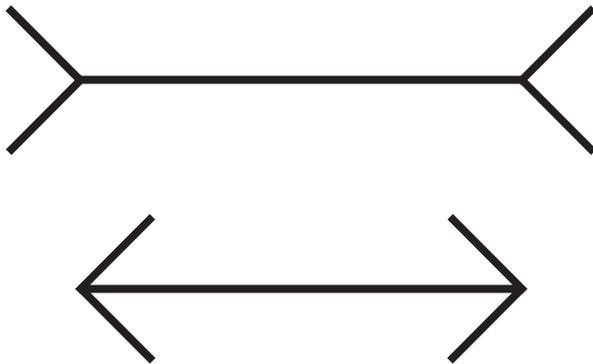


Fig. 12 Müller-Lyer illusion

then it would appear to follow that the other must be as well), but also because I am unconvinced that we truly *see* motion, as opposed to merely *judge* it, in the first place. This is not just a thesis about *illusory* motion, but motion altogether, and will have to be developed and defended in later work. But to give an illustrative example, consider the motion in the rotating dancer illusion: The literature tends to focus on the fact that it is bistable, i.e. that it is liable to switch from clockwise to counterclockwise. But the deeper point is this: we have a 3D rotation in a 2D image; we don't see the dancer as moving *laterally*, but as *rotating*. But we also don't get an impression of the dancer's leg extending beyond the computer screen: there is no stereopsis, it remains a 2D impression. So we paradoxically 'see' motion in a dimension that we do not literally see. A more satisfactory explanation, and one that coheres with my model of pictorial images (see Chap. 3), is that the motion of the dancer is also a post-perceptual unconscious inference rather than something that we see.

Returning to the question of illusory depth, another way of testing my hypothesis is to view the Reverspective monocularly whilst swaying back and forth. Certainly, I can get within 10 cm of the cardboard version I have before the illusion breaks. But at that distance, something weird begins to happen: as I focus on a recess in the Reverspective (that is painted as a protruding building) I get the illusory percept, but I also get the impression of the two physical peaks (that should be the furthest away according to the illusory percept) looming in and out of my vision as I sway back and forth. This is new observation so far as the literature is concerned.

One commentator has suggested to me that this might be an indication that at close distances we experience a mixed percept where *both* the real peaks of the Reverspective *and* the illusory peaks (or real troughs) of the Reverspective are now seen as peaks; leaving the perceived troughs midway down the inverted pyramid, halfway between the real peak and the real trough. But I don't think that this is the correct interpretation. After all, we can place an object (like a pencil) where this illusory trough ought to be and it becomes immediately apparent that the pencil is midway in depth between the real peak and the real trough; just as we would expect from a veridical percept.

But how are we to make sense of the suggestion that it is our *cognition* rather than our *perception* that is being misled by Reverspectives and the hollow-face illusion? Well, consider the question that Wittgenstein posed to Anscombe (recounted in Anscombe 1959):

He once greeted me with the question: ‘Why do people say that it was natural to think that the sun went round the earth rather than that the earth turned on its axis?’ I replied: ‘I suppose, because it looked as if the sun went round the earth.’ ‘Well,’ he asked, ‘what would it have looked like if it had *looked* as if the earth turned on its axis?’

The point being that we don’t persist in thinking there is an *illusion* of the sun going round the earth. Instead, we have come to recognise that the same percept can have two interpretations, one of which may seem more natural but ultimately turns out to be false. Similarly, recall Gregory’s (1970) classic demonstration of the hollow-face illusion of a mask rotating on a stick. I would argue the hollow mask turning from right-to-left looks exactly as it would if it *looked* like a hollow mask turning from right-to-left; the only difference is that we *judge* it to be a protruding mask turning from left-to-right. We might come to this conclusion in two steps: First, I would argue that our visual experience of a 2D video of the hollow-face illusion is entirely consistent with both interpretations, and that we merely *judge* (rather than *perceive*) the mask to have illusory depth and motion. Second, in the case of a real mask, this principle is taken one step further: we judge the mask to have an illusory depth and motion in spite of the fact that, to the extent that stereopsis is present, it points us towards a veridical interpretation. This is just another instance of our *cognition* of a scene or object outstripping our *perception* of it (which, as we have already discussed, may be artefact of our evolutionary need to *interpret* the visual scene as invariant, even though our *perception* of stereopsis falls-off with distance).

But what about the case where we sway back and forth in front of the object: when we have a veridical percept the Reverspective or hollow face appears to remain fixed, but when we have an illusory inverted percept, the Reverspective or hollow face appears to follow us around the room. But again, I would argue that whether something appears fixed or appears to move is not *seen*, but a *post-perceptual inference* applied to what we see. Take the classic case of illusory self-motion: a neighbouring train pulls away from the station, and you mistakenly believe that it is *your* train that is in motion: the veridical interpretation is an equally permissible interpretation of what we see; there is nothing *in our visual experience* that identifies the illusory interpretation over the veridical one.

In conclusion, Gregory (1997) observed that: ‘To maintain that perception is direct, without need of inference or knowledge, Gibson

generally denied the phenomena of illusion'. But in this section I have argued that the denial of *illusions* (as opposed to *delusions*) is not driven by ideological commitments, but by our actual experience: how can we be experiencing an illusory percept if, to the extent we are able to measure the ordinal depth of various points on the Reverspective or the hollow face, they all turn out to be veridical? This account is quite a departure from the contemporary literature where even a purely cognitive explanation of the Müller-Lyer illusion (Fig. 12) is not only assumed to be false, but *obviously* or *self-evidently* false: see Morgan et al. (2012), Witt et al. (2015), and in the Philosophical context Phillips (2016). These articles argue that the tails of the lines in the Müller-Lyer illusion bias our *perception* of their length, whilst I would argue that they merely bias our *post-perceptual evaluation* of their length.

Gregory's work on illusions was influenced by his encounter with Sidney Bradford, a man born blind whose sight was restored following an operation (see Gregory and Wallace 1963; Gregory 2004). Faced with the apparent ineffectiveness of illusions on Sidney Bradford, Gregory concluded that many illusions were acquired over time and must be *cognitive* in nature. Gregory could have drawn one of two implications from this conclusion: either (a) that these illusions were *merely cognitive* rather than perceptual, or (b) that since these illusions were perceptual, perception itself must be cognitive. We all know that Gregory chose the latter interpretation, but is the former really that unsustainable?

4 ATYPICAL RESPONSES TO CUE INTEGRATION

But Sidney Bradford's is not the only atypical response to cue-conflict stimuli:

1. Binocular Depth Inversion Illusion (BDII): If we take a stereogram of a human face and reverse the images, the resulting face will *not* ordinarily be seen as a hollow face by normal observers. This failure of pseudoscopy is often interpreted as pictorial cues vetoing an *unlikely* percept from binocular disparity. By contrast, I would be reluctant to embrace this interpretation before first confirming that the inverted binocular disparity specified a *coherent* depth percept (for instance, simply switching stereo photographs of a real human face—which is the usual stimulus in this context—is liable to introduce discontinuities whenever there is an overhang, and so any 'vetoing' in this context could simply reflect the

fact that no coherent 3D surface can be constructed out of the disparity information). Similarly, there is a question of the extent to which we are merely tracking a *failure* of pseudoscopy (with the eventual percept appearing flat, or at best merely pictorial), rather than a *positive* inversion of depth (as we experience with the hollow-face illusion).

Nonetheless, it has been demonstrated that subjects have a *more accurate* (and *less illusory*) percept of BDII stimuli in the context of (a) schizophrenia (Emrich 1989; Schneider et al. 1996a; Schneider et al. 2002; Koethe et al. 2006, 2009; Dima et al. 2009; Keane et al. 2013, 2016; Gupta et al. 2016), (b) cannabis (Emrich et al. 1991; Leweke et al. 2000; Semple et al. 2003; Koethe et al. 2006), (c) alcohol (Schneider et al. 1998), (d) alcohol withdrawal (Schneider et al. 1996b, 1998), (e) anxiety (Passie et al. 2013), and even (f) sleep deprivation (Schneider et al. 1996a; Sternemann et al. 1997). (Keane et al. 2013, 2016; Gupta et al. 2016 utilise a hollow mask, with which I have no methodological complaint). By contrast, no statistically significant effect was found in the presence of (g) bipolar (Koethe et al. 2009, 2016), (h) dementia (Koethe et al. 2009), (i) depression (Schneider et al. 2002; Koethe et al. 2009), or (j) ketamine (Passie et al. 2003).

But these results appear to pose the following question: just how often in our daily lives would it be detrimental to see the world *more* accurately? For instance, Keane et al.'s (2016) central contention is that those with schizophrenia are able to see the world 'more clearly through psychosis'. But it is difficult to see why this perceptual advantage (specifically, being able to 'more accurately perceive object depth structure') should be problematic? By contrast, being unable to attribute proper context or meaning to whatever we see (my *cognitive* rather than *perceptual* explanation for the failure of depth inversion in this context) would be problematic; leaving those subject to schizophrenia unable to rely on past experience and reliant on ad-hoc rationalisations as they try to make sense of what they see.

2. Child Development: As the case of Sidney Bradford demonstrates, the ability to interpret pictorial cues has to be acquired by experience. For children, this appears to occur around 5–7 months; as Arterberry (2008) explains, by this age children are typically able to rely upon shading, linear perspective, occlusion, texture gradients, familiar size, and surface contours to guide their reaching responses. Depth from disparity also emerges during early child development (around 3–5 months, according to Fox et al. 1980; Held et al. 1980; and Birch et al. 1982).

But as Hong and Park (2008) demonstrate, depth from disparity is relatively coarse for the first 3–4 years (0.68 arc min), and only matures to adult levels of stereoacuity (0.23 arc min) at year 5. Consequently, the early years of child development represent a prime opportunity for Cue Integration to compensate for poor stereoacuity using pictorial cues.

And yet, as Nardini et al. (2010) demonstrate, what is startling is just how late Cue Integration emerges in children. Nardini et al. estimate that Cue Integration of texture and disparity only begins to emerge at around 12 years of age. Certainly, the 6 year olds that they tested did not experience Cue Integration when presented with the cue-conflict stimuli from Hillis et al. (2002). Indeed, the children were able to outperform adults when the sources of information were in conflict, due to an absence of *mandatory fusion*. (And it is worth noting that even for the adults *mandatory fusion* was not complete). So the question is why adults lose access to the single cues relative to children? Given that this loss occurs so late, it doesn't seem plausible that at 12 years of age the processes that govern stereopsis suddenly switches from binocular disparity to Cue Integration. More plausible is the suggestion that experience biases our ability to accurately evaluate the various components of our perception, which would explain the adults' underperformance.

3. Autistic Adolescents: Building on their work with 6 year olds, Bedford et al. (2016) applied the same test to autistic adolescents (12–15 year olds), and found that adolescents with autism integrate cues when it is to their advantage (for instance, two sub-threshold changes in the same direction), but not when it would lead to a reduction in performance (e.g. two opposing changes could be discriminated, rather than cancelling each other out). Bedford et al. suggest that this is a new pattern of behaviour, which they term 'selective fusion', but we've seen this pattern before:

4. Cross-Modal (Vision and Touch): Whilst Hillis et al. (2002) reported mandatory fusion in the context of visual depth cues such as texture and disparity, they came to the apparently paradoxical conclusion that we both *have* and *do not have* perceptual fusion in the cross-modal context of vision and touch:

We also have evidence for *a single, fused percept* for shape information from haptics and vision, *but* in this intermodal case *information from single-cue estimates is not lost*. (emphasis added)

What is going on here? Well, as with Bedford et al., Hillis et al. distinguish between cases where (a) cue combination is likely to lead to an *improved* performance (e.g. two sub-threshold changes in the *same* direction), and (b) cases where cue combination is likely to lead to a *worse* performance (e.g. two above-threshold changes in *opposite* directions), and suggest that subjects only experience Cue Integration when it is to their benefit: i.e. in (a) but not (b).

So, in at least two contexts (autistic adolescents and cross-modal perception), *mandatory fusion* begins to look somewhat less than mandatory. But remember that mandatory fusion (the idea that there are *costs* as well as *benefits* to Cue Integration) was introduced by Hillis et al. (2002) as a means of convincing us that Cue Integration must be operating at the level of *perception* rather than *cognition*. By contrast, the absence of mandatory fusion in (a) autistic adolescents and (b) cross-modal perception appears to suggest that when information is integrated in these contexts it is by virtue of an *integrated judgement* rather than an *integrated percept*. After all, you cannot choose to *see* something as a single, fused percept when it is to your benefit, but as disparate sources of information when it is not. But, and this is the important point, if an *integrated judgement* can explain Cue Integration in the context of (a) autistic adolescents, and (b) cross-modal perception, what makes us so confident that it doesn't explain Cue Integration more broadly?

5. Are we all Atypical? As Oruç et al. (2003) observe, 'large individual differences are the rule in depth perception studies'. But these individual differences are often passed over in the literature without comment: e.g. Hillis et al. (2002), Knill (2007). By contrast, two papers that did make these individual differences a central component were Oruç et al. (2003) and Zalevski et al. (2007). Indeed, the individual differences in Zalevski et al. were so pervasive that: 'No definite conclusion could be drawn ... because of the large variability associated with the estimated weights'. And so, in an ironic twist, the individual differences themselves became the major finding of that study:

The large individual differences we found in cue-combination studies suggest that human observers differ in their cue-weighting strategies and it may be that there is no single model to account for all behaviour, especially when cues to depth are few and in conflict.

Similarly, Oruç et al. (2003) found that although the performance of 6 out of their 8 subjects was consistent with optimal Cue Integration, 2 out of their 8 subjects were positively inconsistent with it. Indeed, these 2 subjects did not even appear to benefit from having two cues rather than one. And it was in light of this, and similar findings in the literature, that Oruç et al. concluded that ‘individual differences abound in depth perception studies (such as the widely ranging cue reliabilities across subjects found by Hillis et al. 2002...)’. What this observation demonstrates is that even the most fervent advocates of Cue Integration (such as Landy and Maloney) recognise that individual differences pose a real concern for their account.

The only alternative is to argue that these individual differences are predicted by Cue Integration itself: for instance, Marty Banks has suggested that Ernst and Banks (2002), Hillis et al. (2004), and Girshick and Banks (2009) demonstrate that different subjects attach different weights to cues based upon the different reliabilities of these cues *for them*. The implication being that although the subjects behave differently, they were all, in fact, exhibiting optimal Cue Integration *given the different reliabilities of each individual cue for each individual subject*. But this raises the question as to why the reliability of each individual cue should vary so drastically between individual subjects (in the way that would be required to account for the results in Oruç et al. 2003; Zalevski et al. 2007)?

But my primary concern is not to pose problems for *optimal* Cue Integration, but to question the level at which Cue Integration occurs in the first place? If Cue Integration really does occur at the level of *perception* then, in light of these pervasive individual differences, we necessarily commit ourselves to the suggestion that even ‘normal’ subjects perceive substantially different slants and angles from one another, with the same scene presented to each observer with an idiosyncratic geometry. By contrast, if Cue Integration merely occurs at the level of *cognition*, then all we have to maintain is that subjects are liable to have idiosyncratic *interpretations* of the very same perceived geometry. The former really is quite a radical conclusion to have to come to in order to accommodate the pervasive individual differences in the literature. By contrast, the latter is exactly what we would expect; namely, that subjects are liable to *interpret* what they see in a variety of different ways.

REFERENCES

- Albertazzi, L., van Tonder, G. J., & Vishwanath, D. (2010). *Perception beyond inference: The informational content of visual processes*. Cambridge, MA: MIT Press.
- 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.
- Anscombe, G. E. M. (1959). *An introduction to Wittgenstein's Tractatus*. London: Hutchinson.
- Arterberry, M. E. (2008). Infants' sensitivity to the depth cue of height-in-the-picture-plane. *Infancy*, 13(5), 544–555.
- Bedford, R., Pellicano, E., Mareschal, D., & Nardini, M. (2016). Flexible integration of visual cues in adolescents with autism spectrum disorder. *Autism Research*, 9(2), 272–281.
- Birch, E. E., Gwiazda, J., & Held, R. (1982). Stereoacuity development for crossed and uncrossed disparities in human infants. *Vision Research*, 22(5), 507–513.
- 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.
- Dima, D., Roiser, J. P., Dietrich, D. E., Bonnemann, C., Lanfermann, H., Emrich, H. M., et al. (2009). Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *NeuroImage*, 46(4), 1180–1186.
- Domini, F., & Caudek, C. (2011). Combining image signals before three-dimensional reconstruction: The intrinsic constraint model of cue integration. In Trommershäuser, Körding, & Landy (Eds.), *Sensory cue integration*. Oxford: Oxford University Press.
- Doorschot, P. C., Kappers, A. M., & Koenderink, J. J. (2001). The combined influence of binocular disparity and shading on pictorial shape. *Perception and Psychophysics*, 63, 1038–1047.
- Eby, D. W., & Braunstein, M. L. (1995). The perceptual flattening of three-dimensional scenes enclosed by a frame. *Perception*, 24(9), 981–993.
- Emrich, H. M. (1989). A three-component-system-hypothesis of psychosis. Impairment of binocular depth inversion as an indicator of functional dysequilibrium. *British Journal of Psychiatry*, 155(S5), 37–39.
- Emrich, H. M., Weber, M. M., Wendl, A., Zihl, J., von Meyer, L., & Hanisch, W. (1991). Reduced binocular depth inversion as an indicator of

- cannabis-induced censorship impairment. *Pharmacology, Biochemistry and Behavior*, 40(3), 689–690.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162–169.
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, 3(1), 69–73.
- Fox, R., Aslin, R. N., Shea, S. L., & Dumais, S. T. (1980). Stereopsis in human infants. *Science*, 207(4428), 323–324.
- Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, 5(11), 1013–1023.
- Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision*, 9(9), 8.
- Glennerster, A., & McKee, S. P. (2004). Sensitivity to depth relief on slanted surfaces. *Journal of Vision*, 4, 378–387.
- Glennerster, A., McKee, S. P., & Birch, M. D. (2002). Evidence for surface-based processing of binocular disparity. *Current Biology*, 12, 825–828.
- Gogel, W. C. (1956). The tendency to see objects as equidistant and its inverse relation to lateral separation. *Psychological Monographs: General and Applied*, 70(4), 1–17.
- Gregory, R. L. (1970). *The intelligent eye*. London: Weidenfeld & Nicolson.
- Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society B*, 352, 1121–1128.
- 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.
- Gupta, T., et al. (2016). Disruptions in neural connectivity associated with reduced susceptibility to a depth inversion illusion in youth at ultra high risk for psychosis. *NeuroImage*, 12, 681–690.
- Hartle, B., & Wilcox, L. M. (2016). Depth magnitude from stereopsis: Assessment techniques and the role of experience. *Vision Research*, 125, 64–75.
- Held, R., Birch, E., & Gwiazda, J. (1980). Stereoacuity of human infants. *Proceedings of the National Academy of Sciences*, 77(9), 5572–5574.
- Held, R. T., Cooper, E. A., & Banks, M. S. (2012a). Blur and Disparity Are Complementary Cues to Depth. *Current Biology*, 22(5), 426–431.

- Held, R. T., Cooper, E. A., & Banks, M. S. (2012b). *Response to Vishwanath*. Originally published alongside Vishwanath (2012a) online, currently unavailable.
- 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*(12), 967–992.
- Hong, S. W., & Park, S. C. (2008). Development of distant stereoacuity in visually normal children as measured by the Frisby-Davis distance stereotest. *British Journal of Ophthalmology*, *92*, 1186–1189.
- Hume, D. (1748). *An Enquiry Concerning Human Understanding*. London: A. Millar.
- Keane, B. P., Silverstein, S. M., Wang, Y., & Pappathomas, T. V. (2013). Reduced depth inversion illusions in schizophrenia are state-specific and occur for multiple object types and viewing conditions. *Journal of Abnormal Psychology*, *122*(2), 506–512.
- Keane, B. P., Silverstein, S. M., Wang, Y., Roché, M. W., & Pappathomas, T. V. (2016). Seeing more clearly through psychosis: Depth inversion illusions are normal in bipolar disorder but reduced in schizophrenia. *Schizophrenia Research*, *176*(2), 485–492.
- 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 Albertazzi, van Tonder, & Vishwanath (Eds.), *Perception beyond inference: The informational content of visual processes*. Cambridge, MA: MIT Press.
- Koethe, D., et al. (2006). Disturbances of visual information processing in early states of psychosis and experimental delta-9-tetrahydrocannabinol altered states of consciousness. *Schizophrenia Research*, *88*(1–3), 142–150.
- Koethe, D., et al. (2009). Binocular depth inversion as a paradigm of reduced visual information processing in prodromal state, antipsychotic-naïve and treated schizophrenia. *European Archives of Psychiatry and Clinical Neuroscience*, *259*, 195.
- Landy, M. S., Maloney, L. T., & Young, M. J. (1991). Psychophysical estimation of the human depth combination rule. In P. S. Schenker (Ed.), *Sensor fusion III: 3-D perception and decognition, Proceedings of the SPIE*, *1383* (pp. 247–254).

- 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.
- Landy, M., Banks, M., & Knill, D. (2011). Ideal-observer models of cue integration. In Trommershäuser, Körding, & Landy (Eds.), *Sensory cue integration*. Oxford: Oxford University Press.
- Leweke, F. M., Schneider, U., Radwana, M., Schmidt, E., & Emrich, H. M. (2000). Different effects of nabilone and cannabidiol on binocular depth inversion in man. *Pharmacology, Biochemistry and Behavior*, *66*(1), 175–181.
- Likova, L. T., & Tyler, C. W. (2003). Peak localization of sparsely sampled luminance patterns is based on interpolated 3D object representations. *Vision Research*, *43*, 2649–2657.
- Maloney, L. T., & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. In W. A. Pearlman (Ed.), *Visual communications and image processing IV, Proceedings of the SPIE*, *1191* (pp. 1154–1163).
- Morgan, M., Dillenburger, B., Raphael, S., & Solomon, J. A. (2012). Observers can voluntarily shift their psychometric functions without losing sensitivity. *Attention, Perception, & Psychophysics*, *74*, 185–193.
- Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences*, *107*(39), 17041–17046.
- Ogline, K. (1959). The theory of stereoscopic vision. In S. Koch (Ed.), *Psychology: A study of a science (vol. I). Sensory, perceptual and physiological formulations* (pp. 362–394). New York: McGraw Hill.
- Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*, 2451–2468.
- Passie, T., Karst, M., Borsutzky, M., Wiese, B., Emrich, H. M., & Schneider, U. (2003). Effects of different subanaesthetic doses of (S)-ketamine on psychopathology and binocular depth inversion in man. *Journal of Psychopharmacology*, *17*(1), 51–56.
- Passie, T., Schneider, U., Borsutzky, M., Breyer, R., Emrich, H. M., Bandelow, B., et al. (2013). Impaired perceptual processing and conceptual cognition in patients with anxiety disorders: A pilot study with the binocular depth inversion paradigm. *Psychology, Health, & Medicine*, *18*(3), 363–374.
- Phillips, I. (2016). Naive realism and the science of (some) illusions. *Philosophical Topics*, *44*(2), 353–380.
- Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision*, *11*(7), 1–17.
- Schneider, U., Leweke, F. M., Sternemann, U., Emrich, H. M., & Weber, M. W. (1996a). Visual 3D illusion: A systems-theoretical approach to psychosis. *European Archives of Psychiatry and Clinical Neuroscience*, *246*(5), 256–260.

- Schneider, U., Leweke, F. M., Niemczyk, W., Sternemann, U., Bevilacqua, M., & Emrich, H. M. (1996b). Impaired binocular depth inversion in patients with alcohol withdrawal. *Journal of Psychiatric Research*, *30*(6), 469–474.
- Schneider, U., Dietrich, D. E., Sternemann, U., Seeland, I., Gielsdorf, D., Huber, T. J., et al. (1998). Reduced binocular depth inversion in patients with alcoholism. *Alcohol and Alcoholism*, *33*(2), 168–172.
- Schneider, U., Borsutzky, M., Seifert, J., Leweke, F. M., Huber, T. J., Rollnik, J. D., et al. (2002). Reduced binocular depth inversion in schizophrenic patients. *Schizophrenia Research*, *53*(1–2), 101–108.
- Sample, D. M., Ramsden, F., & McIntosh, A. M. (2003). Reduced binocular depth inversion in regular cannabis users. *Pharmacology, Biochemistry and Behavior*, *75*(4), 789–793.
- Sternemann, U., Schneider, U., Leweke, F. M., Bevilacqua, C. M., Dietrich, D. E., & Emrich, H. M. (1997). Propsychotic change of binocular depth inversion by sleep deprivation. *Nervenarzt*, *68*(7), 593–596.
- Todd, J. T., & Norman, J. F. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception & Psychophysics*, *65*, 31–47.
- Tyler, C. W. (2004). Theory of texture discrimination of based on higher-order perturbations in individual texture samples. *Vision Research*, *44*(18), 2179–2186.
- Vishwanath, D. (2005). The epistemological status of vision science and its implications for design. *Axiomathes*, *15*(3), 399–486.
- Vishwanath, D., & Domini, F. (2013). Pictorial depth is not statistically optimal. *Journal of Vision*, *13*(9), 613.
- Vishwanath, D., & Hibbard, P. B. (2013). Seeing in 3D with just one eye: Stereopsis without binocular vision. *Psychological Science*, *24*(9), 1673–1685.
- Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, *44*, 289–300.
- 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*, *3*(18), 2685–2696.
- Zalevski, A. M., Henning, G. B., & Hill, N. J. (2007). Cue combination and the effect of horizontal disparity and perspective on stereoacuity. *Spatial Vision*, *20*(1–2), 107–138.



<http://www.springer.com/978-3-319-66292-3>

The Perception and Cognition of Visual Space

Linton, P.

2017, XV, 162 p. 33 illus., Hardcover

ISBN: 978-3-319-66292-3