The Role of Stereopsis, Motion Parallax, Perspective and Angle Polarity in Perceiving 3-D Shape

Aleksandra Sherman 1,∗, Thomas V. Papathomas 2,3,∗, Anshul Jain 4 and Brian P. Keane 3,5

1 Department of Psychology, Visual Perception, Cognition, and Neuroscience Laboratory, Northwestern University, Cresap 301, Evanston, IL 60201, USA
2 Department of Biomedical Engineering, Rutgers University, Piscataway, NJ 08854, USA
3 Laboratory of Vision Research, Center for Cognitive Science, Rutgers University, 152 Frelinghuysen Road, Room A127, Piscataway, NJ 08854-8020, USA
4 Graduate Center for Vision Research, SUNY College of Optometry, New York, NY, USA
5 University Behavioral HealthCare, University of Medicine and Dentistry of New Jersey, Piscataway, NJ, USA

Received 31 August 2010; accepted 9 April 2011

Abstract
We studied how stimulus attributes (angle polarity and perspective) and data-driven signals (motion parallax and binocular disparity) affect recovery of 3-D shape. We used physical stimuli, which consisted of two congruent trapezoids forming a dihedral angle. To study the effects of the stimulus attributes, we used $2 \times 2$ combinations of convex/concave angles and proper/reverse perspective cues. To study the effects of binocular disparity and motion parallax, we used $2 \times 2$ combinations of monocular/binocular viewing with moving/stationary observers. The task was to report the depth of the right vertical edge relative to a fixation point positioned at a different depth. In Experiment 1 observers also had the option of reporting that the right vertical edge and fixation point were at the same depth. However, in Experiment 2, observers were only given two response options: is the right vertical edge in front of/behind the fixation point? We found that across all stimulus configurations, perspective is a stronger cue than angle polarity in recovering 3-D shape; we also confirm the bias to perceive convex compared to concave angles. In terms of data-driven signals, binocular disparity recovered 3-D shape better than motion parallax. Interestingly, motion parallax improved performance for monocular viewing but not for binocular viewing.

© Koninklijke Brill NV, Leiden, 2011

Keywords
3-D shape, stereopsis, motion parallax, perspective, convexity, concavity, illusory motion

∗ To whom correspondence should be addressed. E-mails: aleksandrasherman2014@u.northwestern.edu, papathom@rci.rutgers.edu

© Koninklijke Brill NV, Leiden, 2011 DOI:10.1163/187847511X576802
1. Introduction

Recovery of three-dimensional (3-D) shape from stereo, shading, motion and texture has long been studied in perception. Researchers and artists alike have shown that a variety of pictorial cues such as linear perspective and texture, as well as low-level depth cues, such as binocular disparity and motion parallax, can influence the perception of 3-D objects (Rock, 1983). In the present study, we investigated the influence of various depth cues and the value of motion parallax and binocular disparity, on the recovery of 3-D percepts. In order to assess the role of these cues, we studied certain aspects of a class of art pieces known as reverspectives, invented by Patrick Hughes in 1964 (Slyce, 1998). They are paintings drawn on 3-D surfaces that are usually composed of pointed (or truncated) pyramids — in reverse perspective. Namely, the pyramids ‘jut out’ toward the viewer with the wide bases farther from the viewer than the narrow, truncated top-surfaces. In this way, the painted perspective cues compete against the 3-D surface geometry resulting in bistable 3-D shapes: viewers report perceiving either the veridical structure or the reverse (illusory) depth structure. When viewers perceive the depth illusion, reverspectives appear to move compellingly as viewers move in front of them (Cook et al., 2002; Papathomas, 2002, 2007; Papathomas and Bono, 2004; Rogers and Gyani, 2010; Wade and Hughes, 1999; Wexler, 2003).

Far from being just an artistic curiosity, reverspectives raise important questions on 3-D shape perception that center mainly on the role of top-down influences and on the integration of cues in visual perception and hence can be used to study these processes. Much of the recent evidence suggests a primary role of several cues such as perspective and pictorial cues in recovering the shape of reverspectives (Cook et al., 2002; Papathomas, 2002; Papathomas and Bono, 2004). However, Rogers and Gyani (2010) have shown that, for a wide range of reverspective stimuli, there is little influence of pictorial cues other than perspective cues in recovering the 3-D shape, concluding that it is not necessary to invoke higher-level perceptual processes to explain the illusory motion of reverspectives. Rather, they posit that we perceive the illusory motion as ‘the only real-world scenario (distal stimulus) that could have created the perspective and parallax transformations (proximal stimulus) that stimulate our visual systems’. Reverspectives have also motivated researchers to study how self-motion-elicited parallax cues interact with perspective and other pictorial cues. Cook et al. (2002) suggest that changes in visual information due to self-motion interact with implicit knowledge about anticipated changes, thus eliciting the illusory motion. They explain the motion illusion on the basis of the widening or narrowing of reverspective surfaces that are contrary to expectations for viewers who move in front of the object. Others have also provided explanations as to why reverspectives appear to move, such as Papathomas’s (2007) theory based on earlier work by Gogel (1982, 1990), which posits that the illusory motion is a result of the perceived depth reversal (see also Shimono et al., 2002; Wertheim, 1994). Wexler (2003) further studied visual perception under self-motion and showed that
motion parallax signals are interpreted differently depending on whether they were generated by voluntary head motion or by object motion.

Previous studies have also investigated the interaction of perspective and binocular disparity cues in perception (Knapen and van Ee, 2006; van Ee et al., 2002), the role of pictorial cues (van Ee et al., 2005) and the role of saccades (Both et al., 2003; van Dam and van Ee, 2005) when viewing such ambiguous-depth stimuli. It has been demonstrated that the surface slant of reverspectives is recovered independently by monocular mechanisms that use perspective and binocular mechanisms that use disparity (Knapen and van Ee, 2006). There is also evidence that fast vergence eye movements are governed by the data-driven signals, not the perceived slant (Knapen and van Ee, 2006; Wismeijer et al., 2008). In contrast, Wagner et al. (2008) observed that slow, deliberate vergence eye movements follow the perceived slant, even when it competes with the physical geometry.

Finally, on the neural basis of the illusion, Hayashi et al. (2007) used fMRI techniques, and they speculate that mechanisms involved in mental rotation and depth perception are responsible for the reverspective illusion. Séverac Cauquil et al.’s (2006) findings from VEP recordings also implicate higher cortical areas, particularly in the right hemisphere, for processing perspective depth cues.

Generally, two classes of approaches have been used to study the perception of reverspectives. In one class, researchers employed the entire complex reverspective stimulus, with all its surfaces and pictorial cues (e.g., Cook et al., 2002; Hayashi et al., 2007; Papathomas, 2002; Papathomas and Bono, 2004; Rogers and Gyani, 2010). The other class consists of ‘reductionist’ approaches, in which the stimuli are simple subsets of a reverspective. The simplest such subset is a planar trapezoid that contains linear perspective and binocular disparity cues (Knapen and van Ee, 2006; van Ee et al., 2002; Wexler, 2003). In this paper we use an intermediate approach that enables study of the role of dihedral angles on the illusion by considering a non-planar subset of a reverspective with intermediate complexity: we designed simple 3-D stimuli composed of two trapezoids that form a dihedral angle and contain linear perspective and binocular disparity cues.

One of the main objectives of this study was to examine two key stimulus attributes in recovering 3-D shape (see Note 1): angle polarity (convex or concave) and perspective (proper or reverse); the role of the angle polarity has not been studied for reverspectives. The second main objective was to study the value of motion parallax (Caudek and Proffitt, 1993; Wallach and O’Connell, 1953) and binocular (stereoscopic) disparity (Julesz, 1971; Longuet-Higgins, 1981) in recovering the depth of these 3-D stimuli, and to compare the contribution of these two signals and their interactions (Richards, 1985; Rogers and Graham, 1982; see Todd, 2004, for a review). Even though theoretical analyses have demonstrated that 3-D shape can be recovered accurately under certain conditions for motion parallax (Ullman, 1979) and stereopsis (Longuet-Higgins, 1981), experiments have indicated that humans cannot use all of the available information in these signals to recover depth as accurately as these theoretical analyses would predict (see Todd and Norman, 2003, for
a review). It becomes important, therefore, to study how the human visual system processes motion parallax and disparity to recover 3-D shape. Previous studies have compared the strength of motion parallax and binocular disparity, as well as their interactions, by asking observers to report metric depth relationships (Durgin et al., 1995; Johnston et al., 1994; Rogers and Graham, 1982; Todd and Norman, 2003; van Ee et al., 2005; Wexler, 2003; Wexler et al., 2001). Our approach is very different in that we presented observers with specific bistable physical stimuli that can elicit two competing stable percepts, a veridical (non-illusory depth relationships) and an illusory one (in which depth was reversed); we then assessed the strength of a signal (motion parallax signal, stereopsis signal or their combination) by examining the portion of time (commonly referred to as the predominance) spent in the veridical percept. The term ‘veridical’ applies to the depth of the physical stimulus, not to the one suggested by cues.

We chose to work with physical 3-D stimuli because our work involved stereoscopic viewing and motion parallax produced by self-motion. Previous studies have found differences in perceiving 3-D objects between experiments with physical and computer-simulated stimuli (Buckley and Frisby, 1993; Durgin et al., 1995). Some of the potential sources of these differences are the mismatch between the physical and computer generated self-motion parallax and the difficulty of duplicating the vergence angle and accommodation produced by viewing physical objects. Our choice enables us to include these extraretinal signals that are being used naturally in everyday life, making our results more generalizable and ecologically valid. Further, prior methods have used metric differences in assessing the strength of various cues and their combinations. Our method introduces bistable stimuli that elicit competing percepts and enable observers to respond in such a way that allows us to infer whether they obtain the veridical or the illusory depth percept. Thus, our method affords an additional test that can provide converging evidence, taken together with the results of prior studies, for the role of stimulus attributes and data driven cues, such as binocular disparity and motion parallax.

2. Materials and Methods

We administered two experiments that employed the same stimuli and main procedures. We present the common stimuli and procedures below, and will provide more specific details on how the two experiments differed at the end of this section.

2.1. Participants

A total of 16 naïve observers (18–22 years old) participated in the experiments in exchange for monetary compensation. Participants had normal or corrected-to-normal visual acuity. They had normal stereopsis, as determined by tests with random dot stereograms (Julesz, 1971), and they also had normal color vision, as tested with Ishihara plates (Ishihara, 1917). All observers provided written consent before participating in the study. The experiments were conducted in compliance with the standards set by the IRB at Rutgers University.
2.2. Stimuli

We used mirror-symmetric 3-D stimuli, composed of two congruent trapezoids that formed a convex or concave 90-degree dihedral angle and were connected either along their long or short base (see Fig. 1(A) and Table 1). There were two types of stimuli: (a) Stick stimuli (Fig. 1(B)) were constructed using painted wooden sticks to signal the edges of the trapezoids. (b) Full-surface, or paper, stimuli (Fig. 1(C)), in which the trapezoids were made of cardboard paper.

For each type, we used four stimuli that were combinations of two perspective modes (proper/reverse) and two angle polarities (convex/concave) for a total of eight stimuli (see Table 1). Proper/reverse perspective refers to the distant common vertical edge of the trapezoids being shorter/longer than the near edges: proper perspective indicates that the distant edge is shorter, whereas reverse perspective indicates that the distant edge is longer. Thus proper/reverse perspective provides a depth cue that is consistent with or opposite to the cue provided by binocular disparity. Perspective cues were signaled differently, depending on stimulus type. In stick figures (Fig. 1(B)), they were signaled by the length as well as the width of the sticks: thus, for proper perspective, the near trapezoid vertical edge (base) was longer and thicker than the far base; for reverse perspective, the near base was shorter and thinner than the far base. In both perspective modes the slanted edges

![Figure 1](https://www.brill.nl/sp)
Table 1.
The four stimuli formed by the combinations of two angle polarities (concave/convex) and two perspective modes (proper/reverse)

<table>
<thead>
<tr>
<th></th>
<th>Proper perspective (‘P’)</th>
<th>Reverse perspective (‘R’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concave angle (‘V’)</td>
<td>‘VP’</td>
<td>‘VR’</td>
</tr>
<tr>
<td>Convex angle (‘X’)</td>
<td>‘XP’</td>
<td>‘XR’</td>
</tr>
</tbody>
</table>

that connected the vertical bases were linearly tapered in width. In paper stimuli (Fig. 1(C)), perspective cues were signaled both by the shape of the trapezoids and by a texture gradient: the paper surfaces contained painted texture gradients that conformed to the shape of the trapezoid; thus they provided depth cues that were consistent with the trapezoid’s shape. On the other hand, convex/concave refers to whether the central part of the figure was physically closer to or farther from the observer than the surrounding edges. Within each stimulus type, the dimensions of the convex stimulus and its concave counterpart were selected to project the same monocular frontal retinal image for the viewing distance used in the experiments. The stimuli nominally subtended 5.15° of visual angle vertically and 4.12° of visual angle horizontally. These particular stimulus types were chosen in order to compare the value of pictorial, texture information with simple width and length information in depth judgments.

2.3. General Experimental Approach

We used a chin rest to ensure that observers stayed stationary, when required to do so. To compare the roles of binocular disparity and an equivalent motion parallax, we allowed each observer to move his/her head laterally with an amplitude equal to his/her inter-pupillary distance in both Experiments. We placed vertical ‘stoppers’ to limit the range of head motion. We used four (2 × 2) different combinations of viewing conditions: 2 self-motion modes (stationary (S) or moving (M) observer) × 2 ocular viewing modes (monocular (1) or binocular (2)), resulting in four viewing conditions: S1, S2, M1 and M2. The key comparison — concerning the relative roles of the stereo disparity cue and motion parallax cue in recovering the veridical (see Note 2) depth structure — is between condition S2 (which isolated binocular disparity and excluded motion parallax) and M1 (which isolated motion parallax and excluded binocular disparity). Condition S1 provides a base-
line for performance in the absence of the two cues, whereas condition M2 allows us to study how the two cues combine congruently.

Observers viewed the stimuli from a distance of 200 cm, where the viewing distance was measured from the middle of the object (see diamond symbol in Fig. 1(A)). Durgin et al. (1995) used the same distance for physical stimuli of comparable size. The 200 cm distance was finalized during pilot experiments in which we confirmed that participants could obtain both veridical and illusory depth percepts for most stimuli viewed from 200 cm. In order to reduce the effect of other depth cues on our results, our experiments were conducted in a room with diffuse lighting. The stimuli were placed at the observer’s eye level against a black felt background, to minimize cast shadows that might provide depth cues. Thus, we sought to provide a simple viewing environment so as to focus on the role of the depth cues based on signals that were provided by the objects themselves and not on additional signals of the objects relative to those of the background.

For all of the experiments, we used the Matlab Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) to keep a temporal record of key presses, from which we computed the predominance of the reported depth percepts (namely, the fraction of the total time spent in the corresponding percept). We used the predominance of the veridical depth percept (as inferred by the observers’ responses) as a measure of recovering the true depth structure of the stimuli. Because of reports on the effect of instructions on the perception of multi-stable stimuli (Peterson, 1986; Pitts et al., 2008; Suzuki and Peterson, 2000), we were careful not to bias observers toward any interpretation by instructing them to respond spontaneously to changes in depth percepts.

2.4. Procedure

Each trial consisted of a two-minute viewing of one of the four stimuli in one of the four viewing conditions. Eye dominance was ascertained for each observer using the pointing method (Roth et al., 2002), and the dominant eye was later used in the monocular conditions. Across participants and sessions, we counterbalanced the sequence of the four stimuli presentations and the four viewing conditions to minimize a potential effect of viewing order. A screen hid the stimulus from observers in between trials and it was manually removed at the beginning of each trial. During each two-minute viewing interval, participants were instructed to fixate on a red fixation disk placed at the center of the stimulus and spontaneously report the relative depth of the right vertical edge with respect to the fixation mark.

Experiments 1 and 2 differed in that observers had three and two response options, respectively, to report the depth of the right vertical edge relative to the fixation mark, by pressing appropriate keys on a keyboard (detailed below). In both experiments, each observer participated twice in 16 two-minute trials (4 viewing conditions × 4 stimuli configurations); thus each observer had 64 min of observation time in each experiment.
3. Experiment 1

The paper/stick stimuli formed a between-subject factor where half of the 16 subjects viewed the paper stimuli and the other half viewed stick stimuli. Subjects had three choices for reporting their percepts: they indicated whether the right vertical edge was in front of, behind, or at the same depth as the fixation mark. This three-choice procedure, although atypical, had the advantage of providing a more complete picture of how the brain builds 3-D object structure. For example, if ‘flat’ percepts turn out to be rare, then that would support a ‘winner-take-all’ model, which suggests that only the most salient depth cues (or pre-existing perceptual biases) dominate the final depth percept. By contrast, if flat responses were common, then that would suggest that an approximately flat surface can be represented from depth cues that are individually inconsistent with a flat object. Such an outcome could motivate models that compute object structure from a weighted average of depth cues (see Note 3). Furthermore, we provided the same-depth response to assess how confident observers were in their judgments of the other two choices. In this way, our methodology can yield additional insights into the mechanisms underlying 3-D object perception.

3.1. Results

In this experiment, observers reported the perceived depth structure of the stimulus using three possible answer choices: that the right edge of the stimulus was in front of, at the same depth as, or behind the central fixation. We classified responses as veridical-depth (V) or illusory (reverse)-depth (I) if the percept agreed with or was opposite to the physical geometry, respectively; same-depth (S) responses were kept as such. The data, averaged across 8 observers for each combination of stimulus attributes and viewing conditions, are shown in Table 2. The entries provide the predominance durations for the V and I percepts; of course, the predominance duration of the same-depth percept is $S = 1.0 - (V + I)$. For simplicity of presentation, only significant effects are reported below.

To consider the circumstances that lead to the same-depth (co-planar trapezoids) percept, two mixed-factor ANOVAs were conducted on just the co-planar responses. Stimulus type (paper/stick) was a between-subjects factor in both ANOVAs. Motion and binocularity were within-subject factors in one ANOVA while perspective and angle polarity were within-subjects factors in the other ANOVA. The presence of motion ($F(1, 14) = 23.99, p < 0.001, \eta_p^2 = 0.631$) and binocularity ($F(1, 14) = 13.03, p < 0.01, \eta_p^2 = 0.482$) individually decreased the incidence of co-planar responses, suggesting that fewer depth cues promote flat surface percepts. Furthermore, reverse perspectives produced more flat responses than proper perspectives ($F(1, 14) = 8.32, p < 0.05, \eta_p^2 = 0.373$). This suggests that conflicting depth cues can also lead to more co-planar responses. Third, paper stimuli led to more co-planar responses than stick stimuli ($F(1, 14) = 6.52, p < 0.05, \eta_p^2 = 0.318$), perhaps because the tapering of the stick edges produced better depth
Table 2.
Raw data for Experiment 1. The I and V values represent predominances, i.e., the proportion of time (seconds) spent in the illusory and veridical percepts, respectively.

<table>
<thead>
<tr>
<th>Viewing conditions</th>
<th>Monocular</th>
<th></th>
<th>Binocular</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No motion</td>
<td>Motion</td>
<td>No motion</td>
<td>Motion</td>
</tr>
<tr>
<td>Stick stimuli</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proper perspective</td>
<td>Concave</td>
<td>I = 0.066 (0.034)</td>
<td>I = 0.138 (0.067)</td>
<td>I = 0.028 (0.016)</td>
</tr>
<tr>
<td></td>
<td>Convex</td>
<td>I = 0.035 (0.019)</td>
<td>I = 0.075 (0.035)</td>
<td>I = 0.001 (0.001)</td>
</tr>
<tr>
<td>Reverse perspective</td>
<td>Concave</td>
<td>I = 0.685 (0.069)</td>
<td>I = 0.617 (0.102)</td>
<td>I = 0.350 (0.087)</td>
</tr>
<tr>
<td></td>
<td>Convex</td>
<td>I = 0.712 (0.078)</td>
<td>I = 0.463 (0.077)</td>
<td>I = 0.357 (0.093)</td>
</tr>
<tr>
<td>Paper stimuli</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proper perspective</td>
<td>Concave</td>
<td>I = 0.130 (0.040)</td>
<td>I = 0.098 (0.041)</td>
<td>I = 0.041 (0.027)</td>
</tr>
<tr>
<td></td>
<td>Convex</td>
<td>I = 0.117 (0.060)</td>
<td>I = 0.069 (0.056)</td>
<td>I = 0.090 (0.047)</td>
</tr>
<tr>
<td>Reverse perspective</td>
<td>Concave</td>
<td>I = 0.446 (0.082)</td>
<td>I = 0.374 (0.094)</td>
<td>I = 0.152 (0.083)</td>
</tr>
<tr>
<td></td>
<td>Convex</td>
<td>I = 0.401 (0.102)</td>
<td>I = 0.230 (0.076)</td>
<td>I = 0.161 (0.044)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V = 0.808 (0.078)</td>
<td>V = 0.830 (0.070)</td>
<td>V = 0.934 (0.033)</td>
<td>V = 0.889 (0.062)</td>
</tr>
<tr>
<td></td>
<td>V = 0.723 (0.077)</td>
<td>V = 0.849 (0.060)</td>
<td>V = 0.921 (0.039)</td>
<td>V = 0.887 (0.043)</td>
</tr>
<tr>
<td></td>
<td>V = 0.029 (0.014)</td>
<td>V = 0.194 (0.087)</td>
<td>V = 0.464 (0.106)</td>
<td>V = 0.409 (0.089)</td>
</tr>
<tr>
<td></td>
<td>V = 0.168 (0.064)</td>
<td>V = 0.511 (0.077)</td>
<td>V = 0.593 (0.098)</td>
<td>V = 0.691 (0.065)</td>
</tr>
<tr>
<td></td>
<td>V = 0.504 (0.095)</td>
<td>V = 0.689 (0.081)</td>
<td>V = 0.757 (0.076)</td>
<td>V = 0.862 (0.050)</td>
</tr>
<tr>
<td></td>
<td>V = 0.483 (0.106)</td>
<td>V = 0.712 (0.078)</td>
<td>V = 0.649 (0.097)</td>
<td>V = 0.898 (0.041)</td>
</tr>
<tr>
<td></td>
<td>V = 0.134 (0.069)</td>
<td>V = 0.355 (0.074)</td>
<td>V = 0.520 (0.116)</td>
<td>V = 0.608 (0.083)</td>
</tr>
<tr>
<td></td>
<td>V = 0.125 (0.049)</td>
<td>V = 0.574 (0.073)</td>
<td>V = 0.634 (0.066)</td>
<td>V = 0.756 (0.047)</td>
</tr>
</tbody>
</table>

Standard error values are listed in parentheses.
information than surface texture gradient. Finally, the main effect of perspective interacted with the angle polarity variable ($F(1, 14) = 11.08$, $p < 0.01$, $\eta^2_p = 0.442$).

Specifically, for concave stimuli, a proper perspective produced more flat responses ($p = 0.29$) than a reverse perspective; for convex stimuli, a reverse perspective produced more flat responses ($F(1, 14) = 33.47$, $p < 0.01$, $\eta^2_p = 0.705$). This interaction could be attributed to a convexity bias which neutralizes signals that support a concave percept. These results, taken together, suggest that the conflict, weakness, and paucity of available depth cues can lead to 3-D percepts that are inconsistent with each of those cues (Fig. 2).

Having considered the same-depth responses, we next transformed the three-choice data into a two-choice format. This also allowed us to later make an equitable comparison to the results of Experiment 2, in which observers indeed had two choices. We reasoned that, if observers had been given only two choices in Experiment 1, then the time spent in the same-depth percept for a given set of conditions would be assigned to the two other choices in proportion to their predominance. As detailed in the Appendix, this transformation has the main advantage that it maintains the relative ratios of the two main choices (right vertical edge behind or in front of fixation). We verified that this transformation was judicious by comparing its results to those of the actual two-choice procedure (see results for Experiment 2). We also confirmed that the transformed results did not vary significantly from the original three-choice results by conducting an ANOVA on the original dataset. Thus, we used this new metric and performed mixed factor ANOVAs to assess the roles of perspective and angle polarity and assess the role of viewing conditions.

To assess the role of stimulus attributes, we conducted a mixed ANOVA with perspective (proper/reverse) and angle polarity (convex/concave) as the within-subjects variables and stimulus type as the between-subject variable. We found the predicted significant main effect of perspective ($F(1, 14) = 85.89$, $p < 0.0001$, $\eta^2_p = 0.860$): observers recovered the veridical depth with proper perspective ($M = 0.911$, $t(14) = 9.67$, $p < 0.0001$).

Figure 2. Averages of non-transformed three-choice data for the predominance of the same-depth percept across both stimulus types in Experiment 1. (A) Results for the four stimuli, averaged across all four viewing conditions. (B) Results for the four viewing conditions averaged across all four stimuli. Error bars represent ±1 SEM. Brackets and associated asterisks denote interactions (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).
Figure 3. Averages of transformed two-choice data for the predominance of the veridical percept across both stimulus types in Experiment 1. (A) Results for the four stimuli, averaged across all four viewing conditions. (B) Results for the four viewing conditions averaged across all four stimuli. All error bars represent ±1 SEM. Brackets and associated asterisks denote interactions (*p < 0.05, **p < 0.01, ***p < 0.001).

SEM = 0.021) significantly more often than with reverse perspective (M = 0.516, SEM = 0.038). We also found a significant main effect of angle, F(1, 6) = 16.15, p < 0.0001, $\eta^2_p = 0.920$, in that convex angles (M = 0.755, SEM = 0.027) were more easily recovered than were concave angles (M = 0.671, SEM = 0.022). Furthermore, we observed an interaction between perspective and angle, F(1, 14) = 5.59, p = 0.033, $\eta^2_p = 0.285$ (Fig. 3(A)). Angle polarity had a stronger effect under reverse perspective than under proper perspective. Although there was no main effect of stimulus type (F(1, 14) = 2.65, p = 0.126), the veridical percept was more easily recoverable for paper stimuli (M = 0.749, SEM = 0.031) than for stick stimuli (M = 0.677, SEM = 0.031). We found that there was a significant interaction between stimulus type and perspective suggesting that the observers’ percept was more strongly influenced by perspective cues for stick stimuli as compared to perspective cues in paper stimuli (F(1, 14) = 7.455, p = 0.016, $\eta^2_p = 0.347$).

Next, we evaluated the role of binocular disparity and motion parallax using a mixed ANOVA with ocular viewing (monocular/binocular) and motion (stationary/self motion) as the within-subjects variables, and stimulus type as the between-subjects variable. We found an expected main effect of ocular viewing, where binocular viewing (M = 0.818, SEM = 0.026) produced more veridical percepts than did monocular viewing (M = 0.609, SEM = 0.025), F(1, 14) = 65.65, p < 0.0001, $\eta^2_p = 0.824$. Furthermore, we found a main effect of motion parallax, in which viewing the stimuli while in self-motion (M = 0.760, SEM = 0.028) produced veridical percepts more often than did stationary viewing (M = 0.667, SEM = 0.022), F(1, 14) = 15.44, p = 0.002, $\eta^2_p = 0.525$. Once again, no main effect of stimulus type was observed. A weakly significant interaction between motion and stimulus type (F(1, 14) = 4.99, p = 0.042, $\eta^2_p = 0.263$) indicated that motion was a stronger cue to veridicality for stick stimuli and that motion parallax plays a significant role only for monocular viewing for stick stimuli, (F(1, 3) = 21.629, p < 0.05,
Interestingly, we found an interaction between ocular viewing and motion, $F(1, 14) = 23.28$, $p < 0.0001$, $\eta^2_p = 0.626$ (Fig. 2(B)) suggesting that motion parallax signals were more effective under monocular viewing as compared to binocular viewing condition. Performance was significantly better under the static binocular than the monocular moving condition, indicating that stereoscopic disparity is stronger than motion parallax in recovering the veridical depth percept, in agreement with the results of Durgin et al. (1995). Motion parallax played a significant role in recovering veridical depth for monocular but not for binocular viewing.

4. Experiment 2

The second experiment aimed to concentrate on the two dominant depth percepts (in front of/behind) and confirm the outcomes of the first experiment with a more conventional two-alternative procedure. The subjects for the second experiment were randomly drawn from those who participated in the first. More specifically, four observers were selected randomly from among the eight observers who participated in the corresponding stimulus type (sticks/papers). Thus, each observer experimented with only one type of stimulus throughout the experiments.

4.1. Results

We follow the practice of reporting only on significant effects, as in Experiment 1. In Experiment 2, participants reported their depth percept by pressing one of two keys to indicate whether the right edge was in front of or behind the central fixation mark. As we did for Experiment 1, we report the results in terms of the predominance of the veridical percept. The pattern of results for Experiment 2 is very similar to that of the transformed results in Experiment 1 (Figs 3 and 4). To quantitatively verify the similarity in the results, we performed the same

![Figure 4](image-url)

Figure 4. Averages of two-choice data for the predominance of the veridical percept across stimulus type in Experiment 2. (A) Results for the four stimuli, averaged across all four viewing conditions. (B) Results for the four viewing conditions averaged across all four stimuli. All error bars represent $\pm 1$ SEM. Brackets and associated asterisks denote interactions (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Compare with Fig. 2.
statistical analyses on the results of Experiment 2 as presented above for Experiment 1. First, a mixed factor ANOVA was performed to assess the roles of perspective (proper/reverse) and angle (convex/concave) for both stimulus types. We found the predicted significant main effect of perspective ($F(1, 14) = 69.02$, $p < 0.0001$, $\eta_p^2 = 0.920$): observers recovered the veridical depth with proper perspective ($M = 0.913$, SEM = 0.007) significantly more often than with reverse perspective ($M = 0.562$, SEM = 0.042). We also found a significant main effect of angle, $F(1, 6) = 16.43$, $p = 0.007$, $\eta_p^2 = 0.732$, in that convex angles ($M = 0.770$, SEM = 0.022) were more easily recovered than were concave angles ($M = 0.705$, SEM = 0.024). Furthermore, we observed an interaction between perspective and angle, $F(1, 14) = 15.47$, $p = 0.008$, $\eta_p^2 = 0.721$ (Fig. 4(A)). Again, angle polarity had a stronger effect under reverse perspective than under proper perspective. Figure 5 shows the four combinations of stimulus features across viewing conditions and stimulus type. We found no main effect of stimulus type ($F(1, 14) = 0.911$, $p = 0.377$), nor any significant interactions with stimulus type.

Next, we evaluated the role of binocular disparity and motion parallax using a mixed ANOVA with ocular viewing (monocular/binocular) and motion (stationary/self motion) as the within-subjects variables, and stimulus type as the between-subjects variable. We found an expected main effect of ocular viewing, where binocular viewing ($M = 0.844$, SEM = 0.027) produced more veridical percepts than did monocular viewing ($M = 0.631$, SEM = 0.017), $F(1, 14) = 334.95$, $p < 0.0001$, $\eta_p^2 = 0.982$. Furthermore, we found a main effect of motion parallax, in which viewing the stimuli while in self-motion ($M = 0.784$, SEM = 0.027) pro-

**Figure 5.** Variation of the predominance of the veridical percept with viewing conditions across both stimulus types in Experiment 2. The four panels show this variation separately for each of the four stimulus features. All error bars represent ±1 SEM.
produced veridical percepts more often than did stationary viewing ($M = 0.691$, $SEM = 0.017$), $F(1, 14) = 64.72$, $p < 0.0001$, $\eta^2_p = 0.915$. Likewise, we found an interaction between ocular viewing and motion, $F(1, 14) = 187.55$, $p < 0.0001$, $\eta^2_p = 0.969$ (Fig. 4(B)). Performance was significantly better under the static binocular than the monocular moving condition, indicating that stereoscopic disparity is stronger than motion parallax in recovering the veridical depth percept, in agreement with Experiment 1. Motion parallax played a role in recovering veridical depth for monocular but not for binocular viewing suggesting that the motion module simply ‘goes along with’ the percept obtained by the stereo module. Figure 6 shows the four combinations of viewing conditions across stimulus features and stimulus type. Once again, no main effect and no significant interactions with stimulus type were observed.

In addition to casting the data in terms of the predominance measure, we determined the stability of each percept by indexing the number of reversals. We found few differences in the number of reversals across all conditions, indicating that our stimuli were relatively well matched for stability. However, we did find that in both experiments and for both stimulus types, there was a main effect of perspective ($F(1, 20) = 16.6$, $p < 0.01$). Namely, when the stimulus was in reverse perspective, regardless of angle polarity and viewing condition, there were significantly more reversals, indicating that this percept was more difficult to maintain than its proper perspective counterpart. In light of the predominance results, it is not surprising that there is a significant main effect of perspective — but not of angle polarity.
— on the number of reversals. The predominance data showed that observers were less likely to perceive the veridical structure under perspective changes as compared with angle polarity changes. Thus, because the illusory percept was more common for reverse perspective than it was for concave stimuli, it is plausible that the reverse perspective stimuli elicited less stable percepts, overall, as well.

5. Discussion

In our experiments, we used reverspectives to investigate the role of both stimulus attributes (linear perspective, angle polarity) and perceptual modules (stereo and motion parallax) in recovering the shape of 3-D objects. With respect to stimulus attributes, we employed a $2 \times 2$ design: 3-D objects were composed of two trapezoids that met at a 90-degree dihedral angle (convex or concave); we also manipulated the perspective cue, making it either agree or disagree with the object’s 3-D geometry (proper and reverse perspective, respectively). With respect to perceptual modules, we used four viewing conditions in a $2 \times 2$ design: monocular and binocular viewing, combined with stationary or moving observers. The aim was to measure the efficiency of stereo disparity and motion parallax cues in recovering 3-D structure both independently as well as in conjunction with each other. We used two classes of stimuli to isolate the role of stimulus attributes and perceptual modules: paper stimuli and stick stimuli.

In all the experiments, the pattern of results was very similar. In terms of stimulus attributes, we found that angle polarity did indeed play a significant role in determining the percept. For both classes of stimuli and across all viewing conditions, we found that convex angles produced more veridical responses than concave angles. This result is expected given the bias to view concave surfaces as convex (Hill and Bruce, 1994; Langer and Bülthoff, 2001; Liu and Todd, 2004). This bias for convexity has practical ramifications for the design of reverspectives, as well. Because viewers tend to gaze frequently near the center of a depicted scene, reverspectives that are concave at their centers are expected to elicit a more convincing illusion. This has been tested informally by observing two of Patrick Hughes’s art pieces. In one piece, entitled ‘Photorealism 2009’ (http://www.patrickhughes.co.uk/paintings.htm), the central concave angle elicits a stronger illusion than the central convex angle in another piece, entitled ‘Openings 1999’ (shown as Fig. 2 in Papathomas and Bono (2004)).

Furthermore, we found that the perspective cue helped recover the perceived 3-D structure much more strongly than did angle polarity. Previously, researchers used the motion parallax of random-dot patterns to produce the percept of dihedral angles and reported that humans make more accurate depth order judgments with lower perspective ratios than with higher ones (Braunstein and Tittle, 1988; Braunstein et al., 1993). Thus, it is quite possible that lower perspective ratios might have been even more powerful in eliciting the veridical percept in our experiments.
The experiments also afforded us the opportunity to compare the efficiency of the motion parallax and stereo disparity signals in recovering veridical depth by comparing observers’ performance under two specific viewing conditions: (a) M1 (moving monocular) and (b) S2 (stationary binocular). We found that stereo parallax was significantly stronger in recovering the 3-D shape of the stimuli than motion parallax. This is in agreement with the main findings by Durgin et al. (1995). There is a minor difference, however: they reported that, in pilot studies, motion parallax signals generated by head-movements equal to the inter-pupillary distance were not strong enough to elicit a sense of depth at a viewing distance of 2 m. In contrast, our results from two comprehensive experiments demonstrate that humans can use motion parallax to recover true 3-D shape even at a viewing distance of 2 m, albeit to a smaller degree as compared to depth recovered from stereo cues. The average observed veridical percept predominance was largest for the stereo parallax condition (86%), intermediate for the monocular motion parallax condition (74%) and smallest for the monocular static (S1) condition (53%). However, these averages include stimuli that were designed to elicit illusory percepts. Thus, performance with ordinary objects is expected to be much better, as it is indeed the case.

Interestingly, under binocular viewing, the motion parallax cues did not improve performance; performance was similar for stationary (S2) and moving (M2) binocular conditions. The absence of benefit from the simultaneous presence of binocular disparity and motion parallax has also been suggested by Bradshaw et al. (2000) who similarly found no evidence supporting the hypothesis that the visual system exploits these simultaneous signals. Some of the earlier reports have suggested that the 3-D percept in fact results from an interaction between stereo and motion parallax (Bradley et al., 1995; Johnston et al., 1994; Lankheet and Palmen, 1998; Nawrot and Blake, 1989; Rogers and Graham, 1982). However, Rogers and Collett (1989) reported that motion parallax only affects the perceived depth by stereo parallax for low values of stereo parallax (less than 8′). For these low values, perceived depth was a monotonic function of the motion parallax. Even though the stereo parallax was very small in our stimuli (3.44′), it appears that, unlike their study, the perceived shape was not influenced by the simultaneous presence of motion parallax. It is possible that these differences are due to the different paradigm we used. Observers in Rogers and Collett’s (1989) experiments reported fine depth differences, whereas our observers reported relative depth positions (in front of versus behind).

According to the modified weak fusion model for depth-cue combination (Landy et al., 1995), the motion parallax signal can be used to promote the stereo parallax cue by providing the distance information. Once the stereo cue is promoted, it may be then optimally combined (Yuille and Bulthoff, 1996) with the motion parallax cue according to their respective reliabilities. In our study, the stereo parallax signals were stronger than the motion parallax signals; it is possible that, if we degraded the stereo information by adding noise, the motion parallax could have improved the recovery of 3D shape under binocular viewing. Although previous researchers have argued that stereo parallax plays a secondary role in 3-D shape
perception (Pizlo et al., 2005, 2008, 2010), we find that observers performed best under binocular viewing. Pizlo et al. (2005, 2008) have re-introduced the concept of an abstract 3-D shape as separate from the physical object itself. They posit that 3-D shape perception is governed by the simple shape properties of volume, surface and contours (see Note 1). Their regularization model uses the three shape properties to satisfy four constraints: maximal 3-D symmetry, maximal planarity of contours, maximal 3-D compactness and minimal surface area. In our study, stimuli were very simple and attributes such as symmetry, planarity and compactness were similar among all stimuli. Thus, their differential influence on the percepts was not discernible. Also, the concave percepts invoked by our stimuli lack the property of volume and cannot be handled by the early version of the model (Li et al., 2009). Thus, it is possible that the reason why the illusion is stronger for concave angles is that the percept of a convex angle involves a volumetric object, whereas the percept of a concave angle does not involve volume but only surfaces. Thus it is more natural to perceive a concave angle as convex (resulting in an illusion), but not the other way around.

Additionally, Pizlo et al. (2010) have argued that depth, as well as surfaces, are not the appropriate building blocks for 3-D shape perception. However, the assumption in this paper is that trapezoidal shapes are perceived differently depending on the sign of depth. Indeed, there are previous reports that humans have a tendency to perceive quadrilaterals that have at least one pair of parallel edges as rectangles (Ames, 1951; Griffiths and Zaidi, 2000). Thus, the way observers perceive the shape of these 3-D stimuli may be influenced by this preference for rectangles. This preference correlates with using perspective as a cue for perceiving depth, since the converging edges of a ‘retinal’ trapezoid are usually interpreted as the parallel edges of a rectangle receding in 3-D space.

When the visual system attempts to extract shape from motion, as in the monocular moving condition, there are two extreme possibilities, among others. In one case, it may recover the veridical depth. This is theoretically possible, under the rigidity assumption (Ullman, 1979). Alternatively, it may recover the reverse depth, perhaps due to top-down influences; in this case, the motion parallax signal produces illusory motion and a concomitant percept of non-rigidity. Our results indicate that both of these possibilities materialized, with significant predominance durations of both the veridical and the illusory percepts. The first percept is in agreement with deterministic shape-from-motion models (Ullman, 1979), whereas the second is predicted by probabilistic models that do not assume rigidity (Domini et al., 2006; Tassinari and Domini, 2008). The fact that observers perceive concomitant illusory motion when they obtain a false-depth percept raises a legitimate concern: how does this illusory motion affect their depth judgment? We prepared observers to expect such illusory motion, so as not to be surprised by the unexpected percept. In addition, we instructed them to concentrate on reporting the depth of the right vertical edge relative to the fixation mark and ignore any potential stimulus motion. It is possible that the presence of this illusory motion reduced the effectiveness of the
motion parallax cue. Nevertheless, the difference of the effectiveness of the motion parallax signal between the monocular and the binocular cases could not be explained by such an effect.

It is interesting that, even though observers knew that they were viewing a stable 3-D object, they still perceived drastic changes in depth, switching among the three (or two) choices during the 2-min observation sessions. In particular, the same-depth percept of Experiment 1 is a very peculiar percept because it is supported neither by the data-driven signals, which favored the veridical depth, nor by the perspective cues, which favored either the veridical or the illusory shape, none of which could elicit the same-depth percept. The same-depth percept may have arisen in cases where the top-down and data-driven signals produced competing percepts. Its existence argues against models of the winner-take-all type (Dosher et al., 1986; Lee et al., 1999). Instead, one possibility is that the final percept is a weighted average of the outputs provided by various perceptual modules (Clark and Yuille, 1990; Landy et al., 1995; van Ee et al., 2003). Thus, although the average of the two depth percepts would be unlikely to yield an exactly ‘flat’ percept, the ‘same-depth’ option may have been closest in depth to the final averaged percept. Another possibility is that, whenever neither percept (veridical or illusory) was obtained, then the same-depth percept was reported, only because the option was available to subjects. This suggests that cue conflict may have led to dual discounting and the choice of a neutral description. Generally, the predominance of the same-depth percept is maximal under stationary monocular viewing and its value decreases in viewing conditions that include motion and/or stereo cues. Thus, adding data-driven cues may make it clear to observers that the two trapezoids meet at a near orthogonal dihedral angle and may suppress the alternative interpretation that the two trapezoids may be coplanar.

Interestingly, the same-depth percept was more frequent with the paper stimuli (25% predominance) than with stick figures (10.5% predominance). This increased uncertainty for paper stimuli correlates with our Experiment 1 results in which the effect of perspective was poorer for paper than stick stimuli. This difference between stick and paper stimuli is worth examining; as an anonymous reviewer suggested, it may be due to cognitive influences in that the stick stimuli may appear more three-dimensional than their paper counterparts. In addition, it may result from data-driven cues because texture, being a weaker surface orientation cue, tends to undercut the perception of depth/slant. One potential reason for the observed differences is that, although the ratio of large to small edges in both classes of stimuli was the same, the diameter of the edges co-varied with the perspective cues in stick stimuli and hence might have contributed to the effect. The stronger role that perspective played for stick stimuli than for paper stimuli is in agreement with similar findings of Dosher et al. (1986), who used stick stimuli, in which the luminance of the stick edges varied with the proximity of the edge either congruently or incongruently with disparity-defined depth. They found that this ‘proximity luminance covariance’ was a strong cue to depth. More generally, it is the contrast, rather
than the luminance covariance that provides a cue to depth (Sperling and Dosher, 1995). This variation of contrast with proximity was also present in our stick stimuli because the edge width varied with proximity, causing a concomitant contrast variation, since the white sticks were displayed against a dark background.

The current results add to the growing literature on 3-D shape perception, confirming previous findings such as the unexpected lack of interaction between motion and stereo parallax signals in certain conditions, as well as the role of angle polarity in determining 3-D shape perception. Our study also provides concrete examples where the perceptual system starts with a 3-D model that uses prior knowledge (convexity bias, perspective-driven depth) and progressively modifies it as more and more sensory data (motion parallax, stereo parallax) become available. This is the basis behind numerous Bayesian models for perception (Kersten et al., 2004; Langer and Bulthoff, 2001; van Ee et al., 2003; Weiss et al., 2002), as well as other inferential theories of perception (e.g., Rock, 1983; Wertheim, 1994) that emphasize the interaction of schema-driven and data-driven signals in perception.

Acknowledgements

The authors wish to thank Mr. Tom Grace for invaluable help in building the stimuli and the apparatus used in the experiments. Messrs. Hristiyan Kourtev, Gyuri Schiff and Xiaotao Su provided technical assistance for the projects.

Notes

1. We use the term ‘3-D shape’ rather loosely, to describe how subjects recover one of the two dominant percepts that are possible with our bistable stimuli. According to a model developed to recover 3-D shape with a performance that is remarkably similar to that of humans (Pizlo et al., 2005, 2008, 2010), three shape properties are necessary for the recovery of 3-D shape: volume, surface, and contours. Thus, the 3-D shape of our stimuli is poorly defined or even nonexistent according to the analytical definition of shape in Pizlo et al. (2010), because the concave percepts invoked by the stimuli lack volume (see Section 5).

2. We use the term ‘veridical’ to refer to the depth percept that agrees with the depth relationships in the physical stimulus. For example, a response that the right side is in front of the fixation mark in Fig. 1(A) (and B) would be classified as veridical; the same response given for the stimulus in Fig. 1(C) would be classified as illusory. Notice that Fig. 1(A) is characterized by reverse perspective; namely, the perspective cue elicits a depth relationship that is opposite to the physical depth relationship.

3. It must be noted that cue-averaging is not an optimal strategy for stimuli with cue conflicts as large as the ones used in the current study (Landy et al., 1995).
References


Appendix: Analyzing Experiment 1 by Reassigning the ‘Flat’ Responses

Let us denote by $I$, $V$ and $S$ the predominance durations for the illusory, veridical, and same-depth percepts in Experiment 1; of course, $I + V + S = 1$. We can transform the data of Experiment 1 by trying to cast them as the results of an experiment in which observers had only two choices, as in Experiment 2. A sensible possibility is to assume that, if observers were forced to have only two choices, then, instead of reporting ‘same depth’, they would report ‘in front’ or ‘behind’ with the same relative frequency as they have already exhibited in the relative magnitudes of $I$ and $V$. Thus, we would assign portions of $S$ to $I$ and $V$ in proportion to their values to obtain the transformed values $I'$ and $V'$ as follows:

$$I' = I + \left[\frac{I}{I + V}\right]S, \quad V' = V + \left[\frac{V}{I + V}\right]S.$$  \hspace{1cm} (A.1)

If we expand equation (A.1) we get:

$$I' = \frac{I}{I + V}, \quad V' = \frac{V}{I + V},$$  \hspace{1cm} (A.2)

from which it follows that

$$I' + V' = I + V + S = 1, \quad I'/V' = I/V.$$

(A.3)

Namely, the new measures $I'$ and $V'$ add up to 1, as they should. Furthermore, the magnitudes of ratios $I'/V'$ are maintained at the same level as those of $I/V$ across all conditions, allowing for an equitable analysis in terms of $I'$ and $V'$. 