

# Humans Ignore Motion and Stereo Cues in Favor of a Fictional Stable World

Andrew Glennerster,<sup>1,3,6,\*</sup> Lili Tcheang,<sup>1,4,6</sup>  
Stuart J. Gilson,<sup>1</sup> Andrew W. Fitzgibbon,<sup>2,5</sup>  
and Andrew J. Parker<sup>1</sup>

<sup>1</sup>Department of Physiology, Anatomy and Genetics  
Sherrington Building  
University of Oxford  
Parks Road  
Oxford OX1 3PT  
United Kingdom

<sup>2</sup>Department of Engineering  
University of Oxford  
Oxford OX1 3PJ  
United Kingdom

## Summary

As we move through the world, our eyes acquire a sequence of images. The information from this sequence is sufficient to determine the structure of a three-dimensional scene, up to a scale factor determined by the distance that the eyes have moved [1, 2]. Previous evidence shows that the human visual system accounts for the distance the observer has walked [3, 4] and the separation of the eyes [5–8] when judging the scale, shape, and distance of objects. However, in an immersive virtual-reality environment, observers failed to notice when a scene expanded or contracted, despite having consistent information about scale from both distance walked and binocular vision. This failure led to large errors in judging the size of objects. The pattern of errors cannot be explained by assuming a visual reconstruction of the scene with an incorrect estimate of interocular separation or distance walked. Instead, it is consistent with a Bayesian model of cue integration in which the efficacy of motion and disparity cues is greater at near viewing distances. Our results imply that observers are more willing to adjust their estimate of interocular separation or distance walked than to accept that the scene has changed in size.

## Results and Discussion

In order to study different sources of visual information about the three-dimensional (3D) structure of scenes, it is necessary to bring them under experimental control. Over the past 200 years, a number of ingenious devices and strategies have been used to isolate particular

sources of information so that their influence on human behavior can be assessed (such as Helmholtz's telestereoscope, which effectively increases the separation of the viewer's eyes [5]). A much more general approach is to use the technological advantages of virtual reality to generate a complete visual environment under computer control.

Figure 1 illustrates an observer in a virtual room whose scale varies as the observer walks from one side to the other. Subjects wore a head-mounted display controlled by a computer that received information about the location and orientation of the subject's head and updated the binocular visual displays to create an impression of a virtual 3D environment with a floor, walls, and solid objects. When the virtual room changed size, the center of expansion was half-way between the two eyes (the "cyclopean" point), so that as objects became larger, they also moved farther away. Consequently, no single image could identify whether the observer was in a large or a small room (e.g., images at the top of Figure 1). Thus, the expansion of the room results in retinal flow similar to that experienced by an observer walking through a static room, although the relationship between distance walked and retinal change is altered.

None of the subjects we tested noticed that there had been a change in size of the room. If they reported anything, it was that their strides seemed to be getting longer or shorter as they walked to and fro. The phenomenon is remarkable because binocular and motion cues provide consistent information about the size and distance of objects, and yet the information is apparently ignored. Subjects seem to ignore information both about vergence angle (to overrule stereopsis) and about stride length (to overrule depth from motion parallax).

We tested the consequences of subjects' "blindness" to variations in the scale of the room by asking them to compare the sizes of objects viewed when the room was different sizes. On the left side of the room, the subjects viewed a cube whose size they were to remember. As they walked to the right, the cube disappeared. Then, in a region on the right-hand side of the room, a second cube appeared and subjects were asked to judge whether it was larger or smaller than the first cube. The size of the virtual room varied with the subject's position, as shown in Figure 1. In the period when neither cube was visible, the room expanded gradually until it was four times larger in all dimensions than before. Using a forced-choice paradigm, we determined the size of the comparison cube (viewed when the room was large) that subjects judged to be the same as the size of the standard cube (viewed when the room was small).

Subjects always misestimated the relative sizes of the cubes by at least a factor of 2 and sometimes as much as 4 (see Figure 2). The misestimation varied systematically with the viewing distance of the comparison cube: At far viewing distances, subjects' matches were close to the value predicted if they judged the sizes of the cubes relative to other objects in the room, such

\*Correspondence: a.glennerster@reading.ac.uk

<sup>3</sup>Present address: School of Psychology and Clinical Language Sciences, University of Reading, Reading RG6 6AL, United Kingdom.

<sup>4</sup>Present address: Institute of Neurology, University College London, Queen Square, London WC1N 3BG, United Kingdom.

<sup>5</sup>Present address: Microsoft Research Ltd., 7 JJ Thompson Avenue, Cambridge CB3 0FB, United Kingdom.

<sup>6</sup>These authors contributed equally to this work.

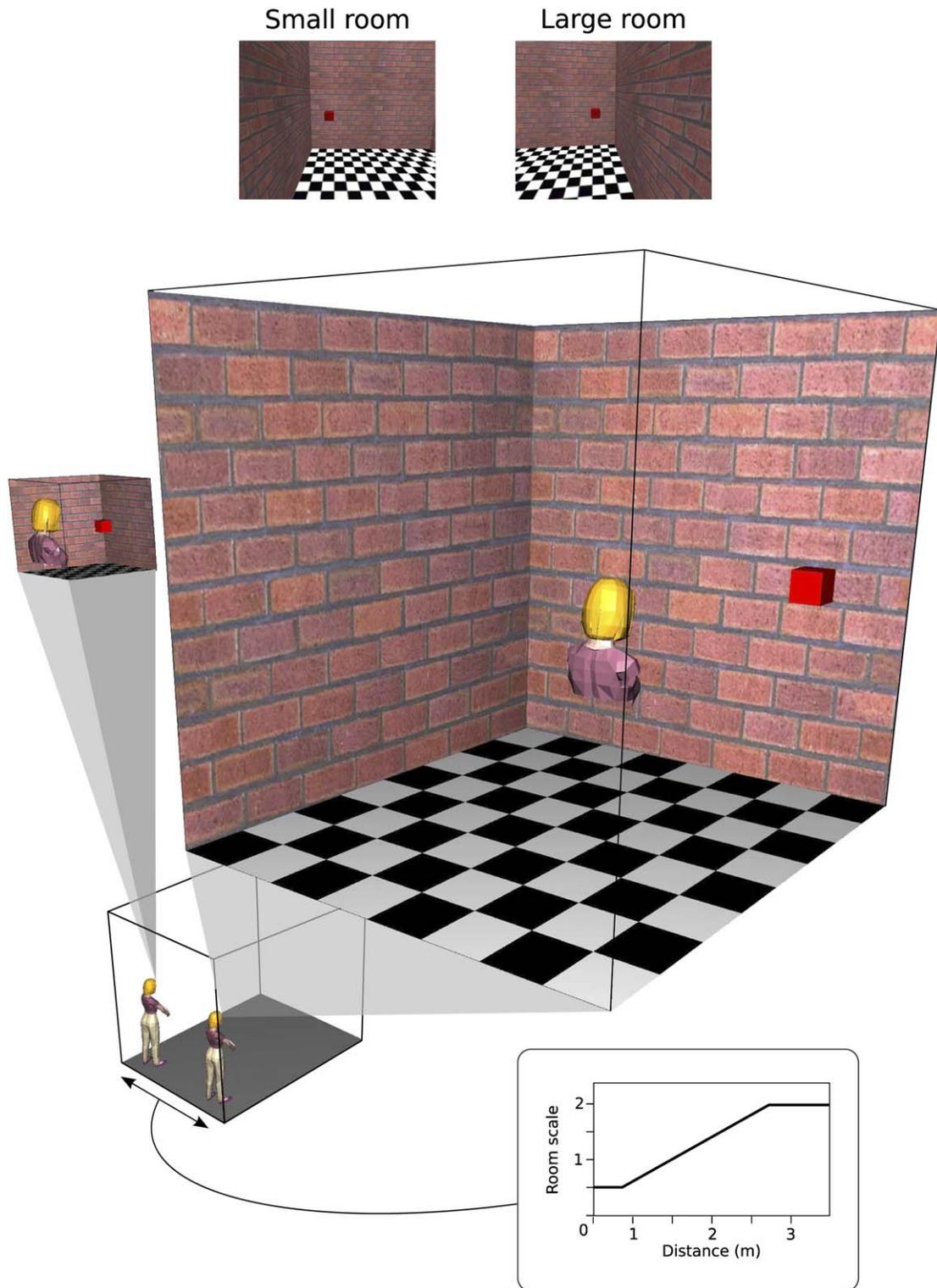


Figure 1. An Expanding Virtual Room

Observers wearing a head-mounted display occupied a virtual room whose size varied as they walked across it. Moving from the left to the right side of the real room caused the virtual room to expand by a factor of 4. The inset graph shows how the scale of the room changed with lateral distance walked. When the scale was 1, the room was 3 m wide and 3.5 m deep. Because the center of expansion was a point midway between the eyes, any single image could not reveal the size of the room, as the example views illustrate. Observers reported no perceived change in the size of the virtual room despite correct and consistent information from stereopsis and motion parallax. In the experiment, observers compared the size of two cubes, one seen when the room was small and the other seen when the room was large.

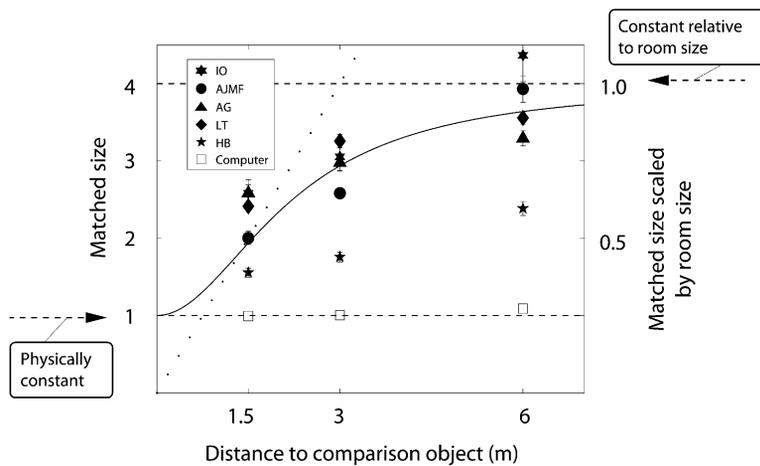


Figure 2. Size Matches in the Expanding Room

The size of the comparison cube that subjects perceived to be the same size as the standard is plotted against the viewing distance of the comparison cube for five subjects. The standard cube was always presented at 0.75 m. The ordinates show matched size relative to the true size of the standard (left) or relative to a cube four times the size of the standard (right), i.e., scaled in proportion to the size of the virtual room. Error bars show  $\pm 1$  standard error of the mean (SEM). The fitted curve shows the output of a model in which cues indicating the true distance of the comparison cube are more reliable, and so have greater weight, at close viewing distances. The dotted line shows the predicted data if subjects matched the retinal size of the standard and comparison cubes. The open squares show the matched size computed from the images a subject saw on a typical trial. We used a 3D reconstruction package, as described in the text and [Supplemental Data](#).

as the bricks forming the wall (a size ratio of 4), whereas at close viewing distances matches were more veridical. An important cue about distance is the height of the eye above the ground plane [9, 10], which does not normally vary in the natural world. In our experiment, the ground plane could provide an important signal for the stability of the room, but the eye-height above the floor varies as the room expands. However, removing the floor and ceiling in the virtual room gives rise to an equally strong subjective impression of stability and similar psychophysical data (see the [Supplemental Data](#) available online).

Stereo and motion parallax, if scaled by interocular separation or distance traveled, should indicate a veridical size match. Hence, it is rational to give more weight to these cues at close viewing distances because this is where they provide more reliable information [11–13]. The curve in [Figure 2](#) shows that a model incorporating these assumptions provides a reasonable account of the data. The single free parameter in the model determines the relative weight given to cues signaling the true distance of the comparison object (e.g., from stereo or motion parallax) compared with cues that specify the size of the cube in relation to the features of the room.

The pattern of errors by human observers is quite different from that predicted by current computational approaches to 3D scene reconstruction. We used a commercially available software package [14, 15] to estimate the 3D structure of the scene and the path that the subject had taken. The input to the algorithm was the sequence of images seen by a subject (monocular input only) on a typical trial, giving separate 3D reconstructions of the room when the standard and comparison cubes were visible. The example in [Figure S2](#) shows the head movement during a typical trial and how the motion-parallax information available to subjects can be used to reconstruct the scene. When combined with information about the actual distance the subject traveled, the algorithm also provides estimates of the size of the room for each sequence of images. The change in room size was recovered almost perfectly,

and hence the computed size matches, shown in [Figure 2](#), are close to 1. If there were errors in the estimate of the distance that the subject traveled, then the size matches for all three comparison distances would have been affected equally, whereas the human data show quite different size matches at different distances.

There is good evidence that information about the distance observers walk and their interocular separation are sufficiently reliable to signal the change in size of the room if these are the only cues. For example, stereo thresholds for detecting a change in relative disparity have a Weber fraction of 10%–20% [16], far below the 4-fold change in relative disparities that occur in the expanding room. Erkelens and Collewijn [17] found, as we do, insensitivity to smooth changes in absolute disparity for a large field stimulus, but in their stimulus, relative disparities did not change. More direct evidence in [Figure 3](#) shows that information from stereo and motion parallax is sufficient to signal the relative size of objects when these cues generate no conflict with texture or eye height information. Subjects initially viewed the standard cube in a small room, the same size as in the first experiment. Instead of the room expanding smoothly as they walked, subjects passed through the wall of the small room into a room that was four times larger, in which the comparison cube was visible. The walls of both rooms were featureless to avoid comparison of texture elements (such as bricks), but stereo- and motion-parallax information was still available from the vertical joints between the back and side walls. The floor and ceiling were also removed to avoid the use of the height of the observer’s eye above the ground as a cue to size of the room [10]. Thus, if observers looked up or down, they appeared to be suspended in an infinite shaft or “well.” [Figure 3](#) (open symbols) shows that size matching across different distances was better than with the smoothly expanding room. The limitation in the expanding room is therefore not due to the lack of motion and disparity information. In fact, in terms of the number of visible contours, there is much less stereo and motion information in this situation. Size matching is even

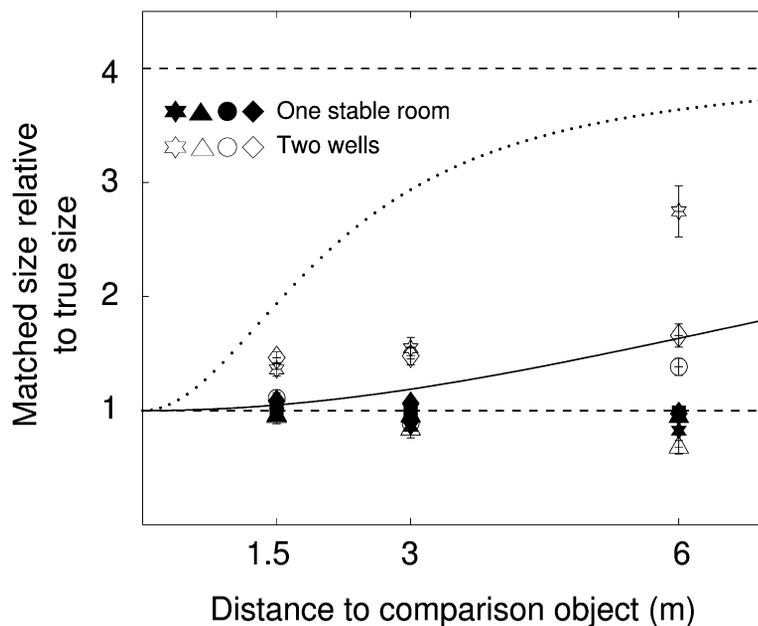


Figure 3. A Static Environment

Size constancy is close to perfect when the comparison task is carried out in a room of constant size (closed symbols, subject symbols as in Figure 2). Size constancy is also significantly improved compared to that in the expanding-room experiment (dotted curve redrawn from Figure 2) when subjects walk from a featureless small “well” (i.e., a room without a floor or ceiling) into a well that is four times the size (open symbols). The solid line shows the fit of equation 5. In this case, stereo- and motion-parallax information dominate. Error bars indicate  $\pm 1$  SEM.

more accurate in a room that remains static (Figure 3, closed symbols), as one would expect from many previous experiments on size constancy [9, 18, 19].

Our results demonstrate that human vision is powerfully dominated by the assumption that an entire scene does not change size. An analogous assumption underlies the classic “Ames room” demonstration [20]. In that illusion, the two sides of a room have different scales but appear similar because observers fail to notice the gradual change in scale across the spatial extent of the room. Our case differs from the “Ames room” illusion because the observers receive additional information about the true 3D structure of the room through image sequences that are rich in binocular disparity and motion information. Nonetheless, the phenomenon is just as compelling. Computer vision algorithms use disparity and motion information continuously to reconstruct the scene (a new reconstruction every moment) [1, 2] (see Supplemental Data). It seems that the human visual system does not. A data-driven process of this kind should signal the current size of the room equally well in the expanding room (Figure 2) or the two wells (Figure 3). Instead, our results are best explained within a Bayesian framework [21] in which a prior assumption that the scene remains a constant size influences the interpretation of 3D cues gathered over the course of many seconds.

#### Experimental Procedures

##### Psychophysics

Subjects (two of the authors and three naive to the purposes of the experiment) viewed a virtual environment by using an nVision data-visor 80 head-mounted display (112° field of view including 32° binocular overlap, pixel size 3.4 arcmin, all peripheral vision obscured). For details of calibration, see [22]. Position and orientation of the head, determined with an InterSense IS900 tracking system, were used to compute the location of the left and right eyes’ optic centers. Images were rendered at 60Hz with a Silicon Graphics Onyx 3200 computer. The temporal lag between tracker movement and corresponding update of the display was 48–50 ms. For the expanding room experiment, the dimensions of the virtual environment varied according to the observer’s location in the real room. When the

observer stood within a zone (0.5 m by 0.5 m, unmarked) near the left side of the room, the size of the virtual room was 1.5 m wide by 1.75 m deep. The standard object, a cube with sides of 5 cm, was always presented 0.75 m from the center of the viewing zone. Subjects were instructed to walk to their right until the comparison cube appeared. They did this rapidly, guided by the edge of a real table (which they could not see in the virtual scene) that ensured they did not advance toward the cubes as they crossed the room. Leaving the first viewing zone caused the standard cube to disappear and the virtual room to start expanding. The center of expansion was the cyclopean point, halfway between the subject’s eyes. The expansion of the room was directly related to the lateral component of the subject’s location between the two viewing zones, as shown in Figure 1. When the scale was 1, the virtual room was 3 m wide and 3.5 m deep. At this scale, the virtual floor was at the same level as the subject’s feet. When subjects reached the viewing zone near the right-hand side of the room, from where the comparison cube could be viewed, the size of the room was 6 m by 7 m. Room size was held constant within each viewing zone. The walls and floor were textured (see Figure 1). There were no other objects in the room. The subject’s task was to judge whether the comparison cube was larger or smaller than the standard, with the comparison-cube size chosen according to a standard staircase procedure [23]. Psychometric functions for three viewing distances of the comparison cube were interleaved within one run of 120 trials. Data from 160 trials per condition were fitted by using probit analysis [24] and the 50% point (point of subjective equality) shown in Figures 2 and 3. Error bars show standard errors of this value, computed from the probit fit. In the two-well experiment, the walls were different shades of gray, and an added black vertical line in each corner meant that the junctions between the back and side walls were clearly visible. These junctions extended without any visible end above and below the observer (as if the observer was in an infinitely deep well).

##### Model

Let  $R$  be the ratio of the size of the comparison object to the size of the standard object and  $\hat{R}$  be the observer’s estimate of this ratio. By definition, when the subject makes a size match in the experiment,  $\hat{R} = 1$ . We consider two different types of cues contributing to  $R$ . We assume one set of “physical” cues (stereo and motion parallax given knowledge of the interocular separation and distance walked) provide an unbiased estimate,  $\hat{P}$ , in other words  $\langle \hat{P} \rangle = R$  where  $\langle \cdot \rangle$  indicates the mean value.  $\hat{T}$  is the estimate provided by cues, such as the texture on the walls and floor, that signal the size of objects relative to the size of the room. The use of texture cues was suggested by Gibson [9]. Because the cubes are not resting on the ground

surface, a cue such as relative disparity is required to identify the texture elements at the same distance as the cube. Because the room expanded 4-fold between the subject viewing the standard and comparison objects, the average estimate of the size of the comparison object according to these “relative” cues is four times smaller (i.e.,  $\langle \hat{T} \rangle = \frac{R}{4}$ ). (As a result, if a subject used only “relative” cues, their match should be four times larger than if they used only “physical” cues.)

If the noises on each of these estimates,  $\hat{P}$  and  $\hat{T}$ , are independent and Gaussian with variances  $\sigma_P$  and  $\sigma_T$  and the Bayesian prior is uniform (all values of  $R$  between 1 and 4 are equally likely a priori), then the maximum-likelihood estimate [12, 13] of the size match is given by:

$$\hat{R} = \hat{P}w_P + \hat{T}w_T = 1 \quad (1)$$

where

$$w_P = \frac{1/\sigma_P^2}{1/\sigma_P^2 + 1/\sigma_T^2}, w_T = \frac{1/\sigma_T^2}{1/\sigma_P^2 + 1/\sigma_T^2} \quad (2)$$

Substituting the average values of  $\hat{P}$ ,  $\hat{T}$ , and  $\hat{R}$  given above into equation 1 and rearranging gives the predicted size match:

$$R = \frac{1}{w_P + w_T/4} \quad (3)$$

We assume that noise on the texture- or room-based size estimate,  $\hat{T}$ , is independent of distance. For example, according to Weber’s law, judging an object relative to the size of neighboring bricks would lead to equal variability at all viewing distances when expressed as a proportion of object size. On the other hand, judging object size by using an estimate of viewing distance introduces greater variability at larger viewing distances. Specifically, under the assumption of constant variability of estimated viewing direction in each eye, the standard deviation of an estimate of viewing distance from vergence increases approximately linearly with viewing distance [11] (see also Figure 12 of [25]). From these assumptions,

$$\frac{\sigma_T}{\sigma_P} = \frac{k}{D^2} \quad (4)$$

where  $D$  is the viewing distance of the comparison object and  $k$  is a constant. From equations 2, 3, and 4, the expected value of the subject’s size match is:

$$R = \frac{k + D^2}{k + D^2/4} \quad (5)$$

Figure 2 shows  $R$  plotted against  $D$ . The curve shows the best fit of equation 5, ( $k = 1.24$ ). The same equation was fitted to the data on two static “wells” shown in Figure 3. In this case, we assume that subjects may still use cues that signal cube size relative to the room, even in the absence of texture. Here,  $k = 33.6$ , indicating a dominance of cues signaling the physical size match.

#### Supplemental Data

Supplemental Data include Supplemental Results and Discussion and two figures and are available with this article online at: <http://www.current-biology.com/cgi/content/full/16/4/428/DC1/>.

#### Acknowledgments

This work was supported by the Wellcome Trust and the Royal Society. We thank O. Braddick, M. Bradshaw, B. Cumming, P. Hibbard, S. Judge, and A. Welchman for critical comments on the manuscript.

Received: November 11, 2005

Revised: January 6, 2006

Accepted: January 9, 2006

Published: February 21, 2006

#### References

1. Faugeras, O.D. (1993). *Three-Dimensional Computer Vision: A Geometric Viewpoint* (Cambridge, USA: MIT Press).

2. Hartley, R., and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision* (Cambridge, UK: Cambridge University Press).

3. Gogel, W.C. (1990). A theory of phenomenal geometry and its applications. *Percept. Psychophys.* *48*, 105–123.

4. Bradshaw, M.F., Parton, A.D., and Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Res.* *40*, 3725–3734.

5. von Helmholtz, H. (1924). *Physiological Optics*, Volume 3 (New York: Dover).

6. Judge, S.J., and Bradford, C.M. (1988). Adaptation to telestereoscopic viewing measured by one-handed ball catching performance. *Perception* *17*, 783–802.

7. Johnston, E.B. (1991). Systematic distortions of shape from stereopsis. *Vision Res.* *31*, 1351–1360.

8. Brenner, E., and van Damme, W.J.M. (1999). Perceived distance, shape and size. *Vision Res.* *39*, 975–986.

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

10. Ooi, T., Wu, B., and He, Z. (2001). Distance determined by the angular declination below the horizon. *Nature* *414*, 197–200.

11. Brenner, E., and Smeets, J.B.J. (2000). Comparing extra-retinal information about distance and direction. *Vision Res.* *40*, 1649–1651.

12. Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* *415*, 429–433.

13. Jacobs, R.A. (2002). What determines visual cue reliability? *Trends Cogn. Sci.* *6*, 345–350.

14. 2d3 Ltd. Boujou 2, 2003. <http://www.2d3.com>.

15. Fitzgibbon, A.W., and Zisserman, A. (1998). Automatic camera recovery for closed or open image sequences. In *Proceedings of the European Conference on Computer Vision*, 311–326.

16. McKee, S.P., Levi, D.M., and Bowne, S.F. (1990). The imprecision of stereopsis. *Vision Res.* *30*, 1763–1779.

17. Erkelens, C.J., and Collewijn, H. (1985). Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Res.* *25*, 583–588.

18. Holway, A.H., and Boring, E.G. (1941). Determinants of apparent visual size with distance variant. *Am. J. Psychol.* *54*, 21–37.

19. Gilinsky, A.S. (1955). The effect of attitude upon the perception of size. *Am. J. Psychol.* *68*, 173–192.

20. Ames, A. (1952). *The Ames Demonstrations in Perception* (New York: Hafner Publishing).

21. Knill, D., and Richards, W. (1996). *Perception as Bayesian Inference* (Cambridge, UK: Cambridge University Press).

22. Tcheang, L., Gilson, S.J., and Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Res.* *44*, 2177–2189.

23. Johnston, E.B., Cumming, B.G., and Parker, A.J. (1993). Integration of depth modules: Stereo and texture. *Vision Res.* *33*, 813–882.

24. Finney, D.J. (1971). *Probit Analysis*, 3rd edition (Cambridge, UK: Cambridge University Press).

25. Hillis, J.M., Watt, S.J., Landy, M.S., and Banks, M.S. (2004). Slant from texture and disparity cues: optimal cue combination. *J. Vis.* *4*, 967–992.