7

Motion-Disparity Interaction and the Scaling of Stereoscopic Disparity

Michael S. Landy
Eli Brenner

Contemporary studies of visual perception often view the observer's problem as one of *estimation*. In other words, the observer seeks to estimate various aspects of the scene, such as the size and shape of the objects in that scene. The information available to inform this estimation is viewed as consisting of one or more visual cues. Each cue may be used on its own to estimate some aspects of scene geometry. Most cues require additional information (visual or otherwise) for the information from that cue to be fully interpreted.

This chapter concentrates on the visual cue to depth of binocular disparity. The cue is the disparity in location of features in the two eyes. Horizontal disparity, the difference in horizontal position in each of the two eyes' views of objects in a scene, is a powerful visual cue to the three-dimensional structure of the world. However, horizontal disparities cannot be fully interpreted without some knowledge of the viewing geometry (the locations, gaze directions, and torsional states of the two eyes). For central gaze (fixating straight ahead), their interpretation requires an estimate of the distance to the fixation point (the *fixation distance*).

Most cues to depth require estimates of the viewing geometry to produce metric estimates of depth in a scene. In a model of depth-cue combination (Landy et al., 1995), this process is described as *cue promotion*. Cues are promoted by the insertion of the values of the unknown viewing geometry and resolution of depth ambiguities. Without promoting the cues, their raw data (e.g., disparities and velocities) are in different units so that simple cue-combination strategies, such as averaging the depth estimates made using each cue, are impossible. When the missing parameters are the eye positions (vergence, gaze directions, and torsions), the promotion process is referred to as *depth scaling*. In particular, in central gaze, the raw sensory data for the cue (velocities, disparities, etc.) are scaled by (that is, multiplied by, or multiplied by the square of) an estimate of the fixation distance. To the extent that this scaling is done accurately, the result is *depth constancy*: perceived depth that is independent of changes in viewing conditions. In this hapter we will limit our discussion of cue promotion to the issue of scaling by the fixation distance.

We review a number of ways in which depth scaling may be accomplished.

We then summarize a series of studies of one such strategy involving combination of horizontal disparity with another depth cue, relative motion, to improve the scaling of horizontal disparity. The addition of motion to a stereo display only improves the interpretation of binocular disparities under very circumscribed conditions: it improves shape perception of the moving object (but no other attributes of the percept of that object) and improves shape perception of nearby objects only if they are very similar (in size, shape, and distance) to the moving object. Thus, it appears unlikely that the interaction between the motion and disparity cues leads to an improvement in the estimate of the fixation distance used to scale disparities and other aspects of the 3-D percept.

7.1 Cue Combination in Depth Perception

Researchers in depth perception describe the information that helps observers estimate depth in terms of individual depth cues such as motion, binocular disparities, vergence, and so on. Each cue can potentially provide independent information concerning the layout of a scene. A thorough listing of such cues numbers well over a dozen (Kaufman, 1974).

Although these depth cues are interesting in their own right, and huge numbers of studies have been done on many of the cues, it is also interesting to understand how observers behave when confronted with multiple cues to depth for the same visual judgment. As we will see, there has been increasing attention given to this problem of cue combination over the years. This problem of combining information from multiple sources has also arisen in computer vision, where it is referred to as the depth fusion or sensor-fusion problem (Aloimonos and Shulman, 1989; Clark and Yuille, 1990). Clark and Yuille (1990) describe different ways in which multiple cues can be combined, ranging from weak fusion, where each cue is used to derive an estimate of depth and then the cues are linearly combined, to strong fusion, which allows for arbitrary nonlinear interactions between the cues.

The problem of combining cues is complicated. Different cues provide different types of information about scene layout. At one extreme is the cue of occlusion, which only gives ordinal depth information at occlusion boundaries. At the other extreme is the cue of vergence, which, at least theoretically, when combined with the other gaze parameters (e.g., version, or gaze azimuth) provides an absolute indication of the distance to the fixated object. In between these extremes lie most of the other cues, which often need to be scaled by an estimate of the fixation distance and/or other viewing parameters to provide metric depth values, and some of which are subject to depth reversals.

The modified weak fusion (MWF) model of depth cue combination Landy et al. (1995) was introduced to describe how a weak-fusion rule (depth-cue averaging) could work for perceived depth. It is based on four principles for depth-cue combination: (1) depth cues are linearly combined using a weighted average of the individual estimates derived from each cue; (2) depth-cue weights are based

on estimates of the cue reliabilities, so that more reliable cues are given greater weight; (3) depth-cue weights are also based on cue consistency and discrepant depth estimates are downweighted, resulting in a robust overall depth estimate; and (4) depth-cues are not averaged until individual cues are promoted to be on the same scale.

The MWF model has been described so far in normative terms (i.e., what characteristics any cue combination rule *should* rationally have. However, some empirical evidence has been gathered for it as a model of human cue combination. A number of researchers have found depth-cue combination to be linear (e.g., Braustein, 1968; Bruno and Cutting, 1988; Cutting and Millard, 1984; Dosher, Sperling and Wurst, 1986; Johnston, Cumming and Parker, 1993; Young, Landy and Maloney, 1993), although others have disputed whether this is really so, while still others even wonder whether 3D shape is determined on the basis of perceived depth at all, rather than on the basis of surface-centered measures such as curvature (Bülthoff and Mallot, 1988; Curran and Johnston, 1994). We have found that cue weights depend on cue reliability (Young et al., 1993). We also have found some evidence for robust cue combination (Li, Maloney and Landy, 1997).

This chapter centers on the issue of cue interaction for cue promotion. Depth cues provide different kinds of information. Object motion as a cue to depth (the *kinetic depth effect*, Wallach and O'Connell, 1953) only provides relative depth information. The actual metric depth of the object must be scaled by an estimate of the fixation distance. Kinetic depth stimuli also undergo depth reversals. Binocular disparities also must be scaled, but in this case depth is approximately proportional to the square of the fixation distance. Thus, the raw data (velocities and disparities) are effectively in different units and until they are scaled by an estimate of the fixation distance, averaging them is a meaningless operation (i.e., the results will depend on the units of measurement chosen). Thus, combinations of these cues rely first on scaling the individual cues using an estimate of the fixation distance.

Some areas of depth vision and cue combination are not easily handled by the preceding model. For example, a number of cues result in stimuli with depth ambiguities. There are regularities in how observers interpret such stimuli, which can be modeled as the result of a priori biases that the observer brings to the depth-interpretation problem. A Bayesian approach to these ambiguities can be brought to bear to estimate the strength and other parameters of such biases (Mamassian and Landy, 1998) as well as how cues with different biases interact to disambiguate a multicue stimulus (Mamassian, Landy and Maloney, in press).

7.2 Depth Scaling

In this section, we review a number of recent studies of depth scaling, with a particular emphasis on the scaling of horizontal disparities and the sources of information used to estimate the fixation distance.

7.2.1 Failures of depth constancy with stereo

A number of researchers have examined whether there is depth constancy for binocular disparities. Wallach and Zuckerman (1963) found good but incomplete constancy with changes of distance when they provided only accommodation and vergence cues to distance consonant with the change in optical distance to the display. Ono and Comerford (1977) describe a number of early studies of depth constancy and review a number of theories. The main result of several studies of the perception of depth from disparity at multiple distances is generally summarized as partial constancy due to a misestimate of the distance (Foley, 1980; Johnston, 1991). Johnston (1991) introduced the *apparently circular cylinder* (ACC) task for examining this issue. In this task, subjects are presented with computer renderings of cylinders with elliptical cross sections, and it is experimentally determined how much depth is required for the cylinders to appear to the subjects to have a circular cross section. For moderately sized, random-dot stereograms, shapes appear distorted in a manner consistent with the hypothesis that observers misjudge the distance, exaggerating it (and perceived depth) when it is substantially less than 1 m, and underestimating it when it is large. Collett, Schwartz, and Sobel (1991) found similar results, with the addition that the size of an object had an impact on scaling; smaller objects were treated as if they were located farther away (and hence were scaled so as to be larger physically and have greater disparity-defined depth).

7.2.2 Distance scaling of size, shape, and depth

Depth is not the only perceptual variable that depends on the viewing geometry. A fixed retinal object should increase in both perceived depth and perceived size (height and width) with an increase in estimated distance. The size should be proportional to the distance and, to first order, the depth proportional to the square of the distance. Hence, aspects of perceived shape (e.g., depth/width, which is the relevant variable in the ACC task) also depend on the viewing distance.

Logically, one might expect the visual system to estimate the viewing geometry as the 6 degrees of freedom of eye position, 3 for each eye, although the binocular extension of Listing's Law implies that there are only 3 degrees of freedom for binocular eye movements toward binocularly fixated targets (see, e.g., van Rijn and van den Berg, 1993), and then use this estimate to scale all items that depend on it (depth, size, shape, etc.). An understanding of scaling necessarily involves several questions. What cues are available to estimate the viewing geometry? How do observers use, weight and combine these cues to estimate the fixation distance and other aspects of viewing geometry? Does this result in a single estimate of viewing geometry that is then used to scale all measurements that require such scaling?

If we restrict ourselves to central gaze, then the only viewing parameter needed is the fixation distance. We will refer to the estimate of the fixation distance used to scale disparities as the *scaling distance*. Several cues to the fixation distance

could conceivably be used. Johnston (1991) found partial constancy in a reduced-cue situation in which vergence and accommodation were the primary cues to distance. There has been some controversy over whether the pattern of vertical disparities between the two eyes' images used for scaling. Helmholtz (1910) was the first to demonstrate the use of vertical disparities using the apparent frontoparallel plane task. Cumming, Johnston and Parker (1991) and Sobel and Collett (1991) concluded that vertical disparities are not used, and Rogers and Bradshaw (1993) and others concluded that they are, with the primary distinction being the size of the display (larger displays yield larger, perhaps more reliable, vertical disparities, Bradshaw, Glennerster, and Rogers, 1996). Another cue involves combining two depth cues that scale differently with the viewing distance (e.g., stereo/motion interaction), a strategy that is the focus of this hapter. Collett, Schwarz, and Sobel (1991) found that relative size affected the scaling distance. Other cues could be used as well, such as the observer's knowledge of the actual distance to the CRTs used in a given experiment or a default value for distance in the absence of other information (*the specific distance tendency* – Gogel and Tietz, 1973).

Given the multiple cues to the fixation distance, its estimation constitutes a cue combination problem to which the principles of MWF might be applied. In our work on stereo/motion interaction, reviewed in Section 7.3.3, we found that the scaling distance was a compromise between that indicated by stereo/motion interaction and that indicated by other cues such as vergence or prior knowledge (Econopouly and Land, 1995). Bradshaw, Glennerster, and Rogers (1996) provide evidence of a weighted combination of vergence and vertical disparity cues to distance, with vertical disparity receiving higher weight with larger displays, which provide larger, more reliable vertical disparities.

The final question is whether these cues, once combined, result in a single value of scaling distance that is then used for all scaling problems that require such an estimate: shape, size, depth, and apparent distance. When one manipulates cues to the fixation distance, there are changes to all of these perceptual attributes, suggesting that there is common distance information used to make all of these perceptual estimates (Rogers and Bradshaw, 1993; van Damme and Brenner, 1997; Brenner and van Damme, 1999).

7.3 Stereomotion Interaction for Depth Scaling

For some time now, we have been examining whether observers combine the depth cues of binocular disparity and object motion to help determine the distance that is then used to scale these cues. The rest of this chapter is concerned with this particular cue interaction. We begin by reviewing why these two cues might be useful in determining the distance. Then, we review the evidence we have gathered as to the circumstances in which it is and is not used.

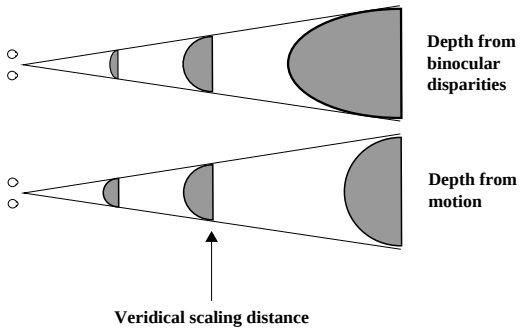

FIGURE 7.1. The scaling problem. Depth from binocular disparities scales approximately with the square of the distance. Thus, a given set of retinal disparities is consistent with a nearby, squashed ellipsoid, a circular cross section at an intermediate distance, or an ellipsoid stretched in depth at a far distance. Depth from relative motion scales linearly with distance. Thus, a motion display consistent with a circular cross section at a nearby distance is also consistent with a larger, but still circular, cross section at farther distances. An interaction of stereo and motion (Richards, 1985; Johnston, Cumming, and Landy, 1994) might involve choosing the distance for which the two cues are consistent, resulting in motion determining the shape and consistency determining the size and distance.

7.3.1 Why combine stereo and motion?

Depth from motion scales linearly with the distance, whereas depth from disparity scales as the square of the fixation distance. Because horizontal extent also scales linearly with the distance, this means that shape from motion (e.g. depth/width) is independent of distance. On the other hand, the same retinal disparities imply different shapes as a function of the distance (Fig. 7.1). Richards (1985) pointed out that one can combine horizontal disparity information with relative motion to derive an estimate of the distance. To do this, one need only select the distance for which the shape estimates from motion and from stereo are in agreement. This cue-interaction scaling hypothesis, were it the only method used for scaling, would result in motion determining the shape (e.g., squashed, circular or stretched in the ACC task). The interaction between motion and disparity would determine the distance and consequently the size.

7.3.2 Evidence with a single object

A number of studies have involved shape judgments with displays that combine binocular disparity and motion cues. Most did not directly address the issue of stereo/motion combination for scaling. For example, Rogers and Collett (1989) added motion parallax to a stereo display of a corrugated surface. The depth percepts from motion alone and disparity alone were in conflict. Observers perceived a curved motion path in the two-cue displays, which, in their stimulus situation, effectively resolved, or at least minimized, the conflict. Tittle and Braunstein (1990) also combined motion and stereo cues, and found evidence for linear cue

combination, with an additional finding of cooperation between the two cues that can be interpreted in terms of motion helping the observers to solve the stereo correspondence problem.

Brenner and van Damme (1999) found that shape judgments improve when object motion is added to a stereo display. In their experiment, subjects adjusted the size and depth of an ellipsoid so that it appeared spherical and the size of a tennis ball they held in one of their hands throughout the experiment. Observers also performed a reach to the apparent distance of the rendered object. Rotating the object improved shape settings, but had little effect on size settings and apparent distance. Thus, the distance-independent shape from motion was used for the shape settings, but the conflict with stereo was not used to improve the distance estimate used to scale size or determine apparent distance. On the other hand, turning the room lights on (to improve cues to distance) improved settings of size, shape, and apparent distance. The results suggest that there are common signals to distance but that the three judgments were handled separately.

The hypothesis that motion and stereo are combined to determine the distance was tested using the ACC task with rotating cylinders (Johnston, Cumming and Landy, 1994). Both stereo and motion were available in some of their displays. They were interested in whether stereo/motion combination, when available, was the sole determinant of the distance estimate. We will show later that stereo/motion may have been used, but only in combination with other estimates of distance.

In their study, observers performed the ACC task for cylinders which either included binocular disparities (versus monocular viewing), structure from motion (rotation back and forth), or both. In the stereo-only condition, settings were biased in a manner that mimicked the previous results of Johnston (1991). The motion-only results were generally accurate, reflecting the lack of distance dependence of structure-from-motion for shape. Settings were also accurate in the condition that included both disparities and motion, consistent with the notion that the two cues are combined in such a way that motion determines the shape (the only perceptual attribute probed by the task), and stereo helps solve for distance and object size. However, in a second experiment, the ACC task was used with stereomotion displays in which the two cues signalled differing amounts of depth (a cue-conflict stimulus). As the depth rendered using disparity was increased, the depth from motion required for an apparently circular cylinder decreased. This implied that binocular disparities did indeed have a weight in the calculation of object shape. It is this apparent contradiction (disparity interpretation is determined completely by motion information, but disparities still have a weight in determining object shape) that led us to do the work described in the next section.

Finally, Tittle, Todd, Perotti, and Norman (1995) found that combinations of binocular disparities and motion did not always result in veridical estimation of shape. They also used the ACC task, but found that the results were different depending on the aspect (the average slant) of the rotating object. They interpreted their results as supporting nonmetric (e.g. affine) representation of shape.

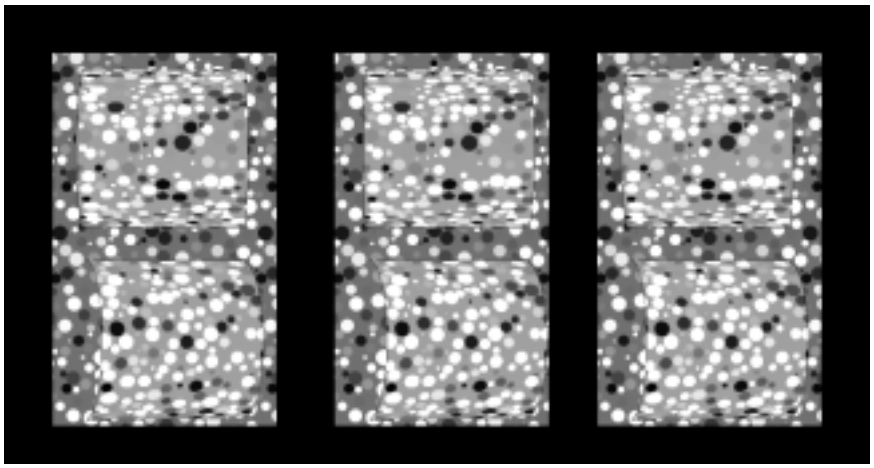

FIGURE 7.2. Example stimulus from a study comparing perceived depth of the top, static cylinder in the context of either a static or rotating bottom cylinder (Econopouly, 1995; Econopouly and Landy, 1995). The left hand image pair is for crossed fusion, and the right hand image pair for diverged fusion.

7.3.3 Two neighboring objects

The results of the study by Johnston, Cumming, and Landy (1994) were puzzling. Their first experiment indicated that the combination of stereo and motion determines the distance used to rescale stereo in a manner intended to make the shape indicated by the two cues consistent. Thus, it appeared that the motion cue determined the perceived shape. The contribution of stereo (in combination with the motion cue), if any, was to determine the distance and size. They did not measure perceived distance and size, and the results of Brenner and van Damme (1999) indicate that stereo/motion combination would not even have had that effect. Because the ACC task only measured perceived shape, then stereo should have had no weight in the results of that task. Direct measurement of cue weights in Johnston et al.'s second experiment revealed a substantial weight for stereo in the combination. How can this be?

One possibility is that motion/stereo combination contributes to scaling but is not the sole source of distance information. We decided to pursue this possibility by adding a second, static object to the rendered scene (Econopouly, 1995; Econopouly and Landy, 1995). The scaling distance used by observers was measured by having them judge the shape of the static, stereo object in the context of a nearby, rotating one. If the interaction between motion and stereo did indeed rescale stereo disparities for these rotating cylinders by helping the observer determine a more accurate estimate of the fixation distance, this improved distance estimate should logically be used to rescale all stereo disparities in the scene. There is, after all, only one fixation distance at a time.

Observers viewed textured, horizontally-oriented cylinders rendered using ray-

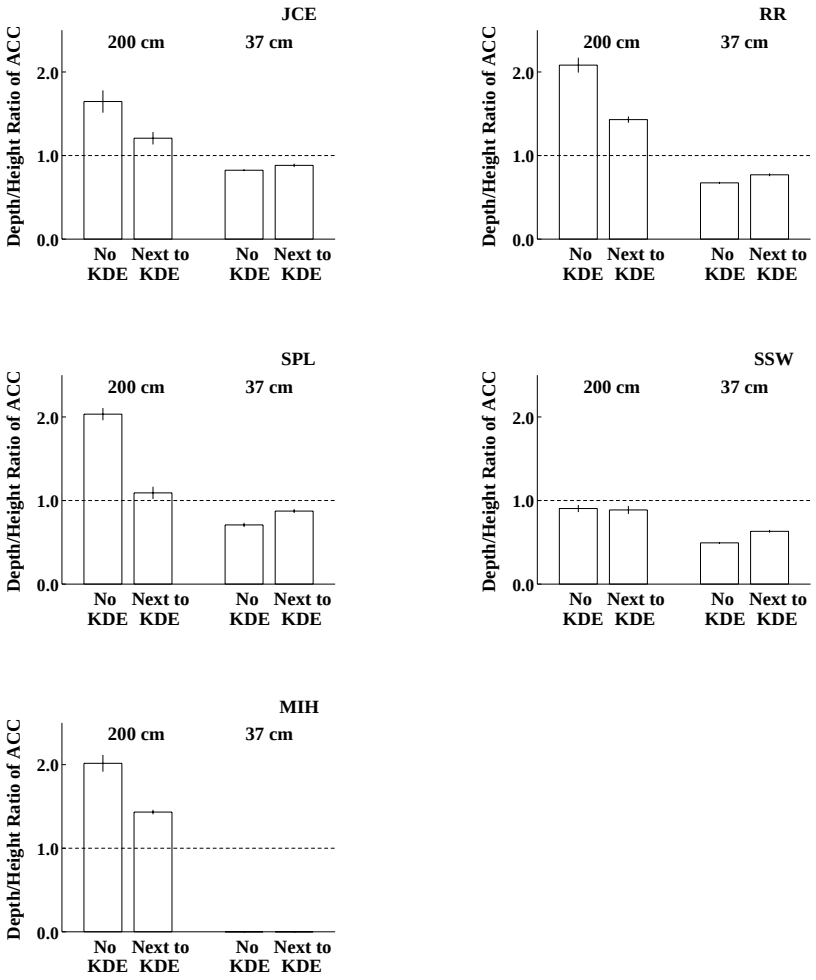

FIGURE 7.3. Apparently circular cylinders (estimated from psychometric functions for forced-choice judgments) for five observers (Econopouly, 1995). For the static cylinders viewed next to a second, static cylinder, the results replicate those of Johnston (1991). In other words, at the 37 cm viewing distance, the depth/height ratio of the ACC was less than 1, implying that observers perceived exaggerated depth (as if they were overestimating the viewing distance when scaling disparity), and at the 200 cm viewing distance the opposite was found. In nearly all cases, the static cylinder was perceived in a more accurate fashion in the context of an adjacent, rotating cylinder (the exception was observer SSW at the 200 cm distance, for which there was no effect of context).

tracing from the optically defined viewpoints of the two eyes (i.e., taking into account the optical path including the stereoscope mirrors). A far (200 cm) and a near (37 cm) distance were used at which, according to the results of Johnston (1991), the distance and hence the depth should be under- and over-estimated, respectively. Two cylinders were shown, one above the other on a static, flat, textured background (Fig. 7.2). The lower cylinder was always rendered with a circular cross section. In one condition, both cylinders were static. In the other condition, the lower cylinder rotated back and forth about a vertical axis. The upper, static cylinder's depth (indicated primarily by binocular disparities, but also by texture and occluding contour) was varied across trials. Observers judged whether the upper cylinder appeared to be stretched or squashed in depth relative to a cylinder with circular cross section (that is, the ACC task). The ACC was estimated as the 50% point on the resulting psychometric function.

The results are clear (Fig. 7.3). When a static cylinder is judged and the adjacent cylinder is also static, binocular disparities are misscaled in a manner that replicates Johnston's findings (1991). When one of the cylinders rotates, not only is it perceived more veridically (Johnston, Cumming, and Landy, 1994), but so is the adjacent, static cylinder. This is completely consistent with the idea that an interaction between motion and disparity in the lower cylinder is used to improve the observer's estimate of the distance, which is in turn used to scale all of the disparities in the scene.

The results of a second experiment (Econopouly, 1995) will allow us to reject several alternative explanations for these context effects. In this experiment, the lower, rotating cylinder was not always rendered as circular in cross section. In Table 7.1, we show the results for conditions in which the rotating cylinder had *consistent cues*. That is, the rotating cylinder was rendered veridically, with binocular disparities and motion both indicating the same amount of depth. The second column shows results when the lower cylinder was static (taken from the previous experiment). They are given in terms of the *effective distance*, which is the distance at which the horizontal disparities of the ACC would be consistent with a rendered circular cylinder. For all three subjects, the effective distances are shorter than the actual viewing distance of 200 cm, which is consistent with the results of Johnston (1991). The third column describes the depth rendered for each cue in the rotating cylinder in units of the cylinder's half-height (so that a circular cylinder corresponds to a value of 1.0). The third column also provides the *equivalent distance*, which is the distance at which the velocities and horizontal disparities are consistent with the same shape as one another (that is, the distance indicated by stereo-motion interaction). For these consistent cue, rotating cylinders, the equivalent distance is the distance for which the stimuli were rendered (i.e., 200 cm). The ACC for the upper, static cylinder was also measured in the context of the rotating cylinder described in the third column whose depth could either be flatter than a circular cylinder (a depth of 0.5), circular (1.0),or exaggerated (1.5 or 2.0). For all three subjects, the upper cylinder ACC setting improved (became more veridical), which can be seen in Table 7.1 as the effective distance becoming closer to the correct value of 200.

TABLE 7.1. ACC results of Econopouly (1995) for static cylinders in the context of rotating cylinders with consistent depth indicated by motion (d_{KDE}) and disparity (d_{BD}) viewed from 200 cm. The values of d_{KDE} and d_{BD} are in units of the cylinder's half-height, so that a value of 1.0 is the depth that would result in a circular cylinder. The second column gives the effective distance for the ACC setting of the upper cylinder when the lower cylinder was static. The third column describes the rendered depth of the lower cylinder. In this table the cues for this cylinder were consistent, so that the equivalent distance was the same as the rendered distance of 200 cm. The fourth column gives the effective distance for the ACC settings of the upper, static cylinder when the lower cylinder was rotating.

Subj.	Effective Distance (No KDE)	Equivalent Distance from BD & KDE		Effective Distance (Next to KDE)
JCE	121 cm	$d_{BD} = 0.5$ $d_{KDE} = 0.5$	200 cm	140 cm
		$d_{BD} = 1.0$ $d_{KDE} = 1.0$	200 cm	145 cm
		$d_{BD} = 1.5$ $d_{KDE} = 1.5$	200 cm	146 cm
		$d_{BD} = 2.0$ $d_{KDE} = 2.0$	200 cm	141 cm
SPL	98 cm	$d_{BD} = 0.5$ $d_{KDE} = 0.5$	200 cm	142 cm
		$d_{BD} = 1.0$ $d_{KDE} = 1.0$	200 cm	148 cm
		$d_{BD} = 1.5$ $d_{KDE} = 1.5$	200 cm	141 cm
		$d_{BD} = 2.0$ $d_{KDE} = 2.0$	200 cm	146 cm
RR	96 cm	$d_{BD} = 0.5$ $d_{KDE} = 0.5$	200 cm	183 cm
		$d_{BD} = 1.0$ $d_{KDE} = 1.0$	200 cm	182 cm
		$d_{BD} = 1.5$ $d_{KDE} = 1.5$	200 cm	172 cm
		$d_{BD} = 2.0$ $d_{KDE} = 2.0$	200 cm	169 cm

There was little or no trend in these ACC settings as a function of the depth of the neighboring, rotating cylinder. This finding argues against depth contrast or assimilation effects. The rotating cylinder can be perceived as flatter or more extended in depth than the upper, static cylinder, and yet its rotation will cause the perceived depth of the upper, static cylinder to increase. In the conditions shown in Table 7.1 it is true, however, that the rotation of the lower cylinder always increases the depth of both the rotating cylinder (Johnston, Cumming, and Landy, 1994) and the upper, static one.

In a second set of conditions (Table 7.2), the lower, rotating cylinder had inconsistent cues. In other words, if interpreted at the optical viewing distance (that indicated by the vergence angle), the depth indicated by motion (and by texture and occluding contour, which always agreed with the motion cue) could be different from that indicated by horizontal disparities. This cue conflict was accomplished by supplying the rendering software with a fallacious value of the inter-pupillary distance (as in Johnston, Cumming, and Landy, 1994), allowing us to exaggerate or diminish the disparities independent of the object motion and texture. You can think of these cue-conflict stimuli as a means of perturbing the equivalent distance (the distance estimate from combining the stereo and motion information). For example, in the second row in the table, depth indicated by motion is exaggerated and depth from disparity is halved (at the rendered distance of 200 cm). This combination of disparities and velocities indicate the same shape if the cylinder is interpreted as located much farther away (an equivalent distance of 554 cm).

The set of rotating cylinders was determined in a control experiment; they were all apparently circular cylinders (this control experiment was a replication of Experiment 2 of Johnston, Cumming, and Landy, 1994). For most subjects and conditions, the effective distance for the static, upper cylinder was a compromise between the effective distance without the context and the equivalent distance from the context (the exceptions are rows 4, 8, and 11 in the table). For subject JCE, the context was able to increase or decrease the effective distance. In other words, the rotation of the lower cylinder could either increase perceived depth, moving the effective distance closer to the veridical value of 200 cm, or decrease depth, making the effective distance less veridical than it already was. This was not the case for the two other subjects, for whom the rotation of the lower cylinder always increased perceived depth of the upper, static cylinder. These results are mostly, but not completely, consistent with the idea that stereomotion interaction results in an estimate of the fixation distance that is then combined (e.g., averaged) with other estimates or defaults before being used to scale disparities. Again, these results argue against any explanation based on the perceived depth of the lower, rotating cylinder affecting perceived depth of the upper cylinder (e.g., contrast or assimilation effects), as a wide range of context effects was achieved using rotating cylinders that were all perceived to have the same circular shape.

Finally, the effect of the rotating cylinder is not due to the motion of the adjacent object per se leading to the rescaling. An adjacent monocular rotating cylinder does not lead to any measurable rescaling.

These results help solve the conundrum of the previous section. Rotating one

TABLE 7.2. ACC results of Econopouly (1995) for static cylinders in the context of rotating cylinders for which depth indicated by motion (d_{KDE}) and disparity (d_{BD}) were inconsistent, viewed from 200 cm. d_{KDE} and d_{BD} are in unit's of the cylinder's half-height; $d_{KDE} = d_{BD} = 1.0$ corresponds to a circular cylinder. The particular values of d_{KDE} and d_{BD} were chosen from a control experiment so that all the rotating cylinders would appear to be circular. Thus, it was hoped that any effect of the context on the upper, static cylinder's appearance would be due to stereo-motion interaction and not simply to the appearance of the lower cylinder (which didn't change across conditions).

Subj.	Effective Distance (No KDE)	Equivalent Distance from BD & KDE		Effective Distance (Next to KDE)
JCE	121 cm	$d_{BD} = 0.0$ $d_{KDE} = 1.4$	∞ cm	207 cm
		$d_{BD} = 0.5$ $d_{KDE} = 1.3$	554 cm	171 cm
		$d_{BD} = 1.0$ $d_{KDE} = 1.1$	221 cm	150 cm
		$d_{BD} = 1.5$ $d_{KDE} = 0.9$	116 cm	128 cm
		$d_{BD} = 2.0$ $d_{KDE} = 0.8$	78 cm	100 cm
SPL	98 cm	$d_{BD} = 0.5$ $d_{KDE} = 1.2$	474 cm	168 cm
		$d_{BD} = 1.0$ $d_{KDE} = 0.8$	164 cm	146 cm
		$d_{BD} = 1.5$ $d_{KDE} = 0.4$	54 cm	134 cm
RR	96 cm	$d_{BD} = 0.5$ $d_{KDE} = 1.2$	488 cm	213 cm
		$d_{BD} = 1.0$ $d_{KDE} = 0.9$	182 cm	175 cm
		$d_{BD} = 1.5$ $d_{KDE} = 0.5$	111 cm	154 cm

cylinder causes the adjacent cylinder to be rescaled so that it appears *more* veridical, but not *completely* veridical. This is true for all of the results in Table 7.1, and 8 of the 11 results in Table 7.2. Thus, the shape derived from disparity is different from that derived from motion, and so any cue-combination rule such as the weighted average we usually find (Landy et al., 1995) could still show a non-zero weight for the disparity cue. This was not seen in the previous data (Johnston, Cumming, and Landy, 1994) because the difference between the stereo-derived shape and the motion-derived shape was less than the variability of the measurements.

7.3.4 Two objects and alternative computations

Thus far, the evidence suggests that stereomotion combination improves the scaling distance. Are there alternative explanations? Suppose that subjects do none of the things we have attributed to them: they do not use stereomotion interaction to estimate the viewing distance, and they do not use this improved distance estimate on either the rotating object or the nearby, static object. Is it possible to account for our results some other way?

Suppose that subjects do not know how to scale disparities. Instead, suppose that their behavior in the various ACC and depth-comparison tasks we have described has been carried soley out by tricks specific to the particular experimental conditions. For a single object (Johnston, Cumming, and Landy, 1994), the rotation might result in a stimulus for which subjects sense cue conflict. They might then set the disparities based on remembered disparities for rotating cylinders they have previously experienced at the given distance. Thus, this disparity correction might *not* result in a new distance estimate and hence should not affect other judgments of this or other objects.

For the case of two objects (Econopouly, 1995), consider a static, stereo cylinder next to a circular, rotating cylinder. In these experiments, the two cylinders lay on a flat, fronto-parallel background, making it abundantly clear that they were located the same distance from the observer. For the results in Fig. 7.3, the rotating cylinder is perceived as approximately circular, therefore the static cylinder, to appear circular, should logically have the same amount of disparity. Thus, observers could have performed the task by, effectively, copying the disparities of the rotating cylinder. The two objects in that study were located at the same distance and had the same width, so that this trick was particularly easy to use (not that subjects were aware of using it).

The required trick is, in fact, slightly more complicated. In Table 7.1, there are conditions in which the rotating cylinder does not have a circular cross section. Thus, we must further posit that an observer viewing, say, a rotating cylinder with twice the depth of a circular cylinder, will halve its disparity before copying it (or comparing it) to the neighboring, static cylinder. For the rotating cylinders containing inconsistent motion and stereo cues (Table 7.2), the rotating cylinder stereo/motion combinations were chosen to result in apparently circular cylinders. The strategy we have described implies that the disparities should have been

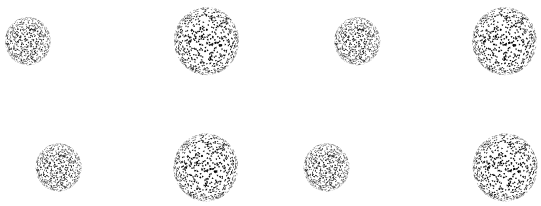

FIGURE 7.4. Example stimulus from Brenner and Landy (1999). The upper image pair is for diverged fusion and the lower one for crossed fusion.

copied to the static cylinder unmodified, resulting in an effective distance (for the static cylinder) equal to the equivalent distance (for the rotating one). This was not found, but the effect of changes in the equivalent distance of the rotating cylinder, although too small, *was* in the appropriate direction.

7.3.5 Two objects at unequal distances

More convincing evidence was needed that observers combined stereo and motion to better estimate the fixation distance. We approached this problem (Brenner and Landy, 1999) by elaborating the experimental paradigm first used by Brenner and van Damme (1999). Observers viewed two textured, stereo ellipsoids, side by side, in an otherwise dark room (Fig. 7.4). The rendered distance to the left-hand ellipsoid was varied over a wide range. Subjects were required to make five adjustments to the rendered objects. First, the left-hand ellipsoid was adjusted to appear equal in size and depth to a tennis ball (i.e., 3.3 cm in radius). Next, the distance to the right-hand ellipsoid was adjusted either to be half that of the fixed, left-hand one, or equal to it (in different blocks of trials). Finally, the size and depth of the right-hand ellipsoid were also set to appear equal that of a tennis ball or, in one condition, to double its size. In some trials, the left-hand ellipsoid rotated back and forth about a horizontal axis; in others, it was stationary. The right-hand ellipsoid was always stationary. If stereo/motion combination resulted in an improved distance estimate, then rotation of the left-hand ellipsoid should have improved the accuracy of all of the settings of *both* ellipsoids.

The results gave little support for stereomotion combination helping observers estimate distance (Fig. 7.5). Individual panels of the figure show one observer's individual settings of width and depth for different conditions. A correct setting for the rendered ellipsoids would have been to adjust all stimuli to lie at the point width = depth = 3.3 cm (except for the double-sized condition). A correct setting of object shape (to be spherical) with the size set incorrectly would have resulted in points along the dashed, diagonal lines. For the static condition (Fig. 7.5a), the settings lie closer to the curve. The points along this curve correspond to disparity and retinal-size settings consistent with a sphere of a fixed size (estimated from the data, larger than the correct value of 3.3 cm for all subjects) located at a

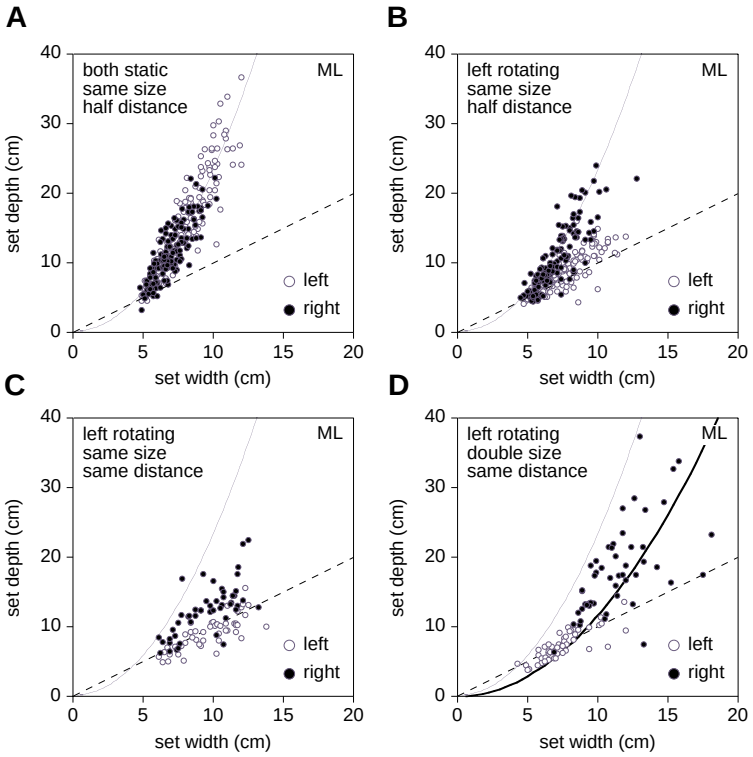

FIGURE 7.5. Results of Brenner and Landy (1999). In each panel, symbols indicate individual settings of half-width and half-depth of the left-hand (open symbols) and right-hand (filled symbols) ellipsoids. Veridical settings would lie at (3.3, 3.3). The dashed line is the locus of true spheres at the rendered distance. The solid curves are the width and depth (at the rendered distance) for which the retinal size and disparity are identical to those of a sphere of a fixed size (not 3.3 cm, but instead for a size determined separately for each subject to fit one condition's data) for a range of distances. It is a curve because the width follows a linear law with distance, and the depth from disparity follows a square law. (a) Both spheres static and the right-hand one set to half the distance of the left-hand one. Settings for both cluster around the curve, consistent with misestimation of the distance. (b) The left-hand sphere is now rotating. Its settings now cluster around the dashed line, and hence are more spherical for the rendered distance. Its size (width) is set no more accurately than before, and settings for the right-hand ellipsoid are not improved. (c) Settings for the case of equal distance. Now both the rotating and static ellipsoids are set to be spherical as in Econopouly (1995). (d) The ellipsoids are set at equal distances, with the right-hand, static one set to be twice the size of a tennis ball. The extra, bold curve corresponds to a double-sized ball at various distances. The scatter is large but appears to follow the bold curve.

distance different than that which was rendered. These results are generally consistent with Johnston (1991). When the left-hand ellipsoid rotates, its settings now cluster about the diagonal line (panel b, open symbols), indicating a setting that is more spherical, consistent with the ACC results of Johnston, Cumming, and Landy (1994). This is also true for settings of the right-hand, static ellipsoid when it is next to a rotating ellipsoid located at the same distance and set to the same size (panel c, solid symbols), consistent with the results of Econopouly (1995).

But, beyond these replications of previous results, all other aspects of the data in Fig. 7.5 are inconsistent with the idea that the rotation of the left-hand, stereo ellipsoid improves observers' estimate of the distance. There is considerable scatter in the size (width) settings of the left-hand, static ellipsoids (Fig. 7.5a, open symbols). These size settings become no less variable or biased when the ellipsoid rotates (panel b, open symbols), which is consistent with the results of Brenner and van Damme (1999). The shape settings of the right-hand ellipsoid do not become more spherical in the context of a rotating ellipsoid when located at half the distance (panel b, filled symbols). These shape settings also improve very little when the distances are equal but the size of the right-hand ellipsoid is set to be double (panel d, filled symbols). For the condition in which subjects halved the distance, the distance settings were unreliable, biased, and unimproved by the rotation of the left-hand ellipsoid (not shown). The set distance was also inconsistent with the set width, which was surprising, as Brenner and van Damme (1999) found that subjects can copy, double, or halve distances across changes in version (gaze azimuth) when only vergence is available, even though they can not judge absolute distance well at all. Finally, subjects were poor at the size settings across the halved distance, even though all they needed to do was to set the retinal size of the right-hand ellipsoid to be double that of the left-hand one.

To summarize, although we replicated the previous findings of improved shape settings for rotating, stereo objects and for static, stereo objects located next to them, these effects are very restricted. The improvement of shape settings for the neighboring, static object only occurred if it was located at the same distance and had the same size as the rotating object. Also, improvements in shape settings were not associated with improvements in set size or set distance. Thus, there is very little evidence for an underlying, improved estimate of the distance used to scale the various scene attributes that logically require distance scaling.

We remained puzzled by this since the trick we suggested to explain Econopouly's results does not fully explain them, and subjects are certainly not aware of using such a trick. We ran another experiment (Landy and Brenner, 1999) to try to determine the limits of the effect of context: How similar in size and distance must the two objects have to be for the rotation of one to result in improved shape settings for the other? The methods and stimuli were identical, in most respects, to the previous experiment (Brenner and Landy, 1999). Again, two ellipsoids were rendered side by side. The right-hand ellipsoid was static; the left-hand ellipsoid was static in half the trials and rotating back and forth about a horizontal axis in the other half of the trials. The distance of each was chosen from three possible values (225, 300 or 400 cm), as was the size of each (12.6, 16.8 or 22.4 cm

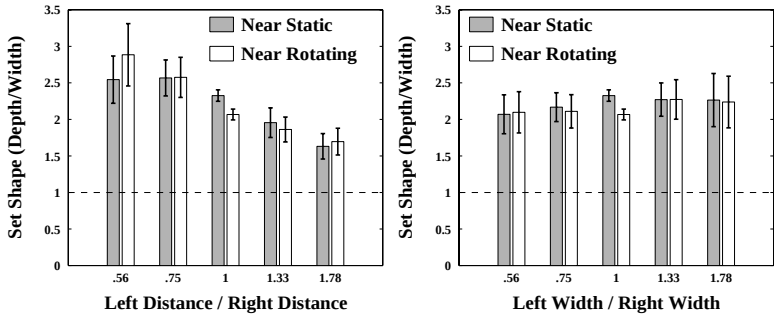

FIGURE 7.6. Shape settings from the study of Landy and Brenner (1999). A value of 1.0 indicates a veridical, spherical setting. Error bars indicate plus or minus two standard errors. Settings for the static, right-hand ellipsoid only became more veridical in the context of a rotating, left-hand ellipsoid when the two ellipsoids were located at the same distance (left panel) and had the same size (right panel).

diam). Observers adjusted the depth of each ellipsoid so that it appeared spherical (thus, there were two adjustments to make per trial). Six observers ran 368 trials each. In about half the trials, the two ellipsoids were the same size and at the same distance. In the other half, the size and distance were chosen at random. As in the previous experiment (see Fig. 7.5, Brenner and Landy, 1999), subjects almost always set the depth too large. Rotating the left-hand ellipsoid reduced this error for the rotating ellipsoid. Figure 7.6 summarizes the shape settings (set depth divided by rendered width) for the other, right-hand ellipsoid. A veridical setting results in a set shape value of 1.0. Again, there was no significant improvement of the shape setting of the static, right-hand ellipsoid in the context of a rotating, stereo ellipsoid unless the two ellipsoids were located at the same distance (left panel; left/right = 1) and had the same size (right panel; left/right = 1). Note that the smaller standard errors in these conditions are due to the large number of trials in which distance and width are identical for the two ellipsoids.

The question remains: When, if ever, does stereomotion combination result in improved scaling, and what object attributes' scaling is improved by it? It is possible that all of the effects described here stem from associations and heuristics as described in Section 7.3.4. It is also possible that the scaling distance is improved, but this improved estimate is only applied under very limited circumstances.

In all of the experiments described here, eye movements were unconstrained. Typically, observers would fixate one object and then the other, changing vergence if they lay at different distances. Perhaps the stereo-motion interaction only applies while the rotating object is actively fixated, or only across iso-vergent saccades from that rotating object. When one judges a second, static object lying at a different distance (Brenner and Landy, 1999), then the improved scaling distance from the rotating object must be combined with information about the required vergence change to fixate the static object to estimate the latter's distance. In Brenner and Landy (1999), the distance was halved, and both objects were never seen

fused at the same time. In the final study (Landy and Brenner, 1999), the differences in distance were substantially smaller and the objects were often seen fused simultaneously. Nevertheless, observers changed fixation from object to object to make their settings. In fact, observers find it nearly impossible to do these tasks when asked to maintain fixation on a spot lying off of the object to be judged.

7.4 Summary

It would be logical for observers to estimate scene attributes such as object size, depth, shape, and distance using all the information available to them. For most of these attributes, as estimated by many available cues, complete estimation requires the observer to promote the measurements (relative velocities, shading gradient, binocular disparities, etc.) by parameters related to the viewing geometry. When stereo is the only cue to depth available, shape is often misestimated as if the wrong value of the viewing distance were being used. Shape estimates improve when an object rotates, or when an object is next to an identical, rotating object. But it seems that these are the only estimates of scene attributes that are improved by the presence of a rotating, stereo object. Although the information is available to refine an observer's estimate of the distance using the concurrent stereo and motion information, observers either do not use it at all, or use it only to improve shape estimates under very restricted circumstances. Viewing distance estimates may well be computed using multiple cues (e.g., Bradshaw, Glennerster, and Rogers, 1996), but our evidence suggests that the combination of stereo and motion rarely, if ever, contributes to such estimates.

Acknowledgments

This work was supported by NIH grant EY08266 and the AFOSR. We thank Larry Maloney and Marty Banks for comments on the manuscript.

References

Aloimonos, J., and Shulman, D. (1989). *Integration of Visual Modules: An Extension of the Marr paradigm*. New York: Academic Press.

Bradshaw, M. F., Glennerster, A., and Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vis. Res.*, 36:1255–1264.

Braunstein, M. L. (1968). Motion and texture as sources of slant information. *J. Exp. Psych.*, 78:247–253.

Brenner, E., and van Damme, W. J. M. (1998). Judging distance from ocular convergence. *Vis. Res.*, 38:493–498.

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

Brenner, E., and Landy, M. S. (1999). Interaction between the perceived shape of two objects. *Vis. Res.*, 39:3834–3848.

Bruno, N., and Cutting, J. E. (1988). Minimodularity and the perception of layout. *J. Exp. Psych. Gen.*, 117:161–170.

Bülthoff, H. H., and Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *J. Opt. Soc. Am. A*, 5:1749–1758.

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

Collett, T. C., Schwarz, U., and Sobel, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, 20:733–754.

Cumming, B. G., Johnston, E. B., and Parker, A. J. (1991). Vertical disparities and perception of three-dimensional shape. *Nature* , 349:411–413.

Curran, W., and Johnston, A. (1994). Integration of shading and texture cues: Testing the linear model. *Vis. Res.*, 34:1863–1874.

Cutting, J. E., and Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *J. Exp. Psych. Gen.*, 113:198–216.

van Damme, W., and Brenner, E. (1997). The distance used for scaling disparities is the same as the one used for scaling retinal size. *Vis. Res.*, 37:757–764.

Dosher, B. A., Sperling, G., and Wurst, S. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vis. Res.*, 26:973–990.

Econopouly, J. (1995). Binocular disparities and kde are combined to rescale binocular disparities. Unpublished doctoral dissertation.

Econopouly, J. C., and Landy, M. S. (1995). Stereo and motion combined rescale stereo. *Invest. Ophthalmol. Vis. Sci. Suppl.*, 36:S665.

Foley, J. M. (1980). Binocular distance perception. *Psych. Rev.*, 87:411–434.

Gogel, W. G., and Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Percept. Psychophys.*, 13:284–292.

von Helmholtz, H. (1910/1925). *Treatise on Physiological Optics*. New York: Dover.

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

Johnston, E. B., Cumming, B. G., and Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vis. Res.*, 34:2259–2275.

Johnston, E. B., Cumming, B. G., and Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture. *Vis. Res.*, 33:813–826.

Kaufman, L. (1974). *Sight and Mind*. New York: Oxford.

Landy, M. S., and Brenner, E. (1999). When does motion added to one object improve the judged shape of a nearby, static object? *Invest. Ophthal. Vis. Sci. Suppl.*, 40:S801.

Landy, M. S., Maloney, L. T., Johnston, E. B., and Young, M. J. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vis. Res.*, 35:389–412.

Li, J. R., Maloney, L. T., and Landy, M. S. (1997). Combination of consistent and inconsistent depth cues. *Invest. Ophthal. Vis. Sci. Suppl.*, 38:S903.

Mamassian, P., and Landy, M. S. (1998). Observer biases in the 3d interpretation of line drawings. *Vis. Res.*, 38:2817–2832.

Mamassian, P., Landy, M. S., and Maloney, L. T. (in press). Bayesian modeling of visual perception. In R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki (Eds.), *Statistical Models of the Brain*. Cambridge, MA: MIT Press.

Ono, H., and Comerford, J. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and Constancy in Visual Perception: Mechanisms and Processes* (pp. 91–128). New York: Wiley.

Richards, W. (1985). Structure from stereo and motion. *J. Opt. Soc. Am. A*, 2:343–349.

van Rijn, L. J., and van den Berg, A. V. (1993). Binocular eye orientation during fixations: Listing's law extended to include eye vergence. *Vis. Res.*, 33:691–708.

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

Rogers, B. J., and Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quart. J. Exp. Psych.*, 41A:697–717.

Sobel, E. C., and Collett, T. S. (1991). Does vertical disparity scale the perception of stereoscopic depth? *Proc. Roy. Soc. (Lond.) B*, 244:87–90.

Tittle, J. S., and Braunstein, M. L. (1990). Shape perception from binocular disparity and structure-from-motion. In P. S. Schenker (Ed.), *Sensor Fusion III: 3-D Perception and Recognition*, Volume 1383, (pp. 225–234).

Tittle, J. S., Todd, J. T., Perotti, V. J., and Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *J. Exp. Psych.: Human Percept. and Perf.*, 21:663–678.

Wallach, H., and O'Connell, D. N. (1953). The kinetic depth effect. *J. Exp. Psych.*, 45:205–217.

Wallach, H., and Zuckerman, C. (1963). The constancy of stereoscopic depth. *Am. J. Psych.*, 76:404–412.

Young, M. J., Landy, M. S., and Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vis. Res.*, 33:2685–2696.