

## 14 Size and speed constancy

---

*Suzanne P. McKee and Harvey S. Smallman*

In his 1963 *Nature* paper, Richard Gregory defined size constancy as “the tendency for objects to appear much the same size over a wide range of distances in spite of the changes of the retinal images associated with distance of the object.” As this definition makes clear, size constancy is about the appearance of objects, about what things look like. Strictly speaking, size constancy denotes only that the apparent size of an object is nearly invariant with changes in distance, not that the observer perceives the true physical size of an object. This invariance implies, however, that some process corrects the angle subtended on the retina by some measure of relative distance, and thus that observers have good information about the *relative* physical size of the objects surrounding them. If the body is used to provide a “metric” for both size and distance, such as a hand viewed at arm’s length, then, in principle, true physical size could be estimated with some degree of accuracy (Morgan, 1989). In this chapter, we will examine how well observers estimate objective size.<sup>1</sup> Because speed constancy is often treated as an extension of size constancy, we will also look at the human ability to estimate objective speed.

### Size constancy

Stripped of phenomenology, size constancy seems to be a fairly simple problem in visual processing. The human visual system measures the angle subtended by an object on the retina, estimates the object’s relative distance, and scales the measured retinal subtense by the estimated distance to obtain an estimate of objective size (Andrews, 1964; Boring, 1946; Epstein, 1973). In a computational context that stresses measurement and scaling, the argument between the Structuralist and Gestalt schools becomes a question of information access. Does the observer have direct access to information about retinal subtense, as the Structuralists might claim (left side of Figure 14.1)? Or is the information about

This chapter was supported by AFOSR Grant F49620-95-1-0265, NEI Grant RO1-EY06644, and NEI Core Grant EY06883. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the National Eye Institute, or the U.S. government.

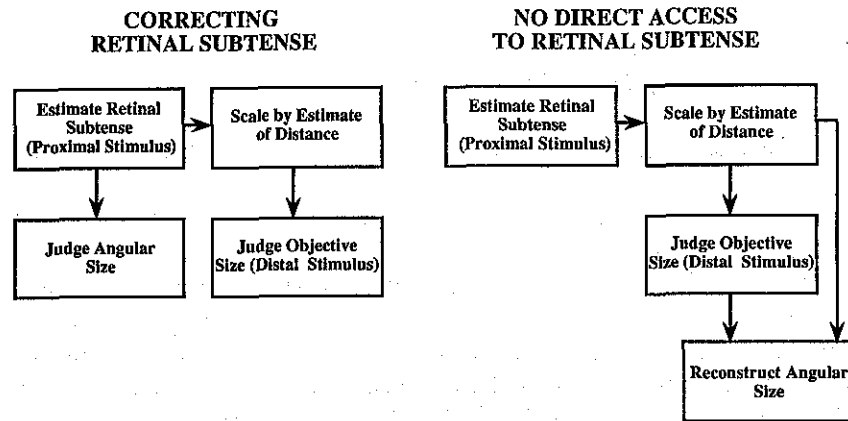


Figure 14.1. Flow charts showing two different conceptual frameworks for size constancy.

retinal subtense lost in the scaling process, so that the observer must compensate for constancy scaling in order to estimate angular size indirectly, as the Gestaltists would claim (right side of Figure 14.1)?

Although the flow charts convert a philosophical controversy into an explicit computational formulation, the representation of size processing in these diagrams is simplistic. For one thing, it is unlikely that normal observers have direct access to information about *retinal* subtense per se. Obviously, all human judgments involve cortical processing, but some types of processing can alter the retinal geometry, converting it into a form that is no longer isomorphic with the angles subtended on the retinae. An example of this type of conversion is the lateral separation between two features at different viewing distances, for example, telephone poles at different distances from the head. The apparent separation between them does not correspond to the lateral separation on either of the two retinae because the visual direction of the feature lying off the fixation plane is roughly halfway between the disparate locations on the two retinae. Does this fact about binocular processing imply that the Gestalt sequence shown on the right of Figure 14.1 is correct? Not necessarily. Angular information is exceptionally useful, so it is difficult to believe that all information about angular subtense is lost or discarded in the processing of size and is recalculated only in response to the artificial demands of a psychophysical experiment. For example, a systematic change in the angles subtended by the elements of a regular pattern (*texture gradient*) is interpreted as a change in relative depth, indicating that angular measurements are accessible at cortical levels responsible for the processing of depth and shape. Moreover, we are certainly aware of these texture gradients, despite perceptual evidence that the receding pattern is composed of

elements of the same size. Is this angular information known directly (left side of Figure 14.1) or indirectly (right side of Figure 14.1)? How can we tell?

The most serious problem with the sequences shown in Figure 14.1 is that no experimental test would decide between them. Both sequences could produce veridical estimates of objective and angular size, provided that the initial estimates of retinal subtense and distance to the object were correct and the observer interpreted the experimental instructions appropriately. It might seem that the indirect estimate of angular size (Figure 14.1, right) would be less exact than the direct estimate (Figure 14.1, left) because it involves additional operations, but that is true only if the additional operations introduce additional uncertainty or noise into the judgment. What is missing from the sequences in Figure 14.1 are the pervasive sources of biological noise that limit all human judgments. Even the most attentive observer will make mistakes in judging objective size because of noise in the sensory processes that encode the relevant dimensions. These errors in discriminating between stimuli, or in matching one stimulus to another, provide clues about the coding sequence that underlies human size judgments. Rather than concentrating on what is perceived, we examine how well human observers can judge objective size, no matter what their percepts are. What kinds of errors do they make in judging either angular or objective size? We use the results from numerous perceptual and psychophysical studies to construct a better diagram of how the human brain processes both angular and objective size.

#### *Accuracy and precision*

Although the terms *accuracy* and *precision* are often used interchangeably, they refer to different types of measurement errors (Bevington, 1969). Accuracy indicates how close a given measurement comes to the true value, whereas precision shows the reliability (or variance) of the measurement. These two indices of error are independent. Random noise affects the precision of human judgments but not their accuracy. A measurement can be accurate but very imprecise, as when the mean of a set of judgments equals the true value but has a large standard error. A measurement can also be precisely wrong (inaccurate), as in the case of a systematic bias with a small standard error.

Accuracy and precision are easily distinguished in psychophysical studies. For example, in size discrimination experiments, observers are asked to judge whether the test stimulus is larger or smaller than the standard. The percentage of trials on which the test is judged larger is plotted as a function of its physical size, and the resulting psychometric function can be fitted with a cumulative normal curve. The cumulative normal curve has two independent parameters: the mean, which defines the location of the curve along the stimulus axis, and the standard deviation, which defines the steepness of the curve. The mean of

the fitted function corresponds to the *point of subjective equality (PSE)*, which refers to the value of the test stimulus that the observer sees as equal to the standard. The slope of the function determines the precision of the increment threshold; one common definition of threshold is the incremental change in the stimulus that produces one standard deviation change in response ( $d' = 1.0$ ). The most useful estimate of precision is the dimensionless Weber fraction (threshold/mean) because it allows comparison of the biological errors associated with different dimensions, such as size errors with disparity errors.

For the most part, studies of constancy have focused only on accuracy – on how well observers reproduce (*match*) the objective size of the test object. Usually, judgments from all observers are lumped together to estimate a common mean and a standard deviation. This approach necessarily confounds the diversity of the sample with the variability of individual performance. For example, observers tend to overestimate an object's size with increasing distance, a phenomenon known as *overconstancy* (Carlson, 1960; Gilinsky, 1955; Sedgwick, 1986). However, the standard deviation of the pooled judgments from many observers often overlaps the correct physical size of the test object. Does this result mean that all observers see the test object as larger than its true size, but that there is considerable variability in their judgments? Or does it mean that some observers consistently see the test object as larger, whereas a smaller number consistently encode the correct size of the test object? Ideally, constancy studies should examine human diversity (many observers), as well as assess individual precision (repeated measurements on the same observer).

Few studies have made direct measurements of the Weber fractions for objective size, but precision can be estimated indirectly by dividing the average error or standard deviation of an individual's judgments by his or her mean. The flow charts in Figure 14.1 suggest that there are at least two sources of potential noise in any objective size calculation: (1) the noise in the estimate of the angle subtended on the retina and (2) the noise in the estimate of the object's distance. The noise in angular subtense is best inferred from the increment threshold for lateral separation between targets presented in the fixation plane, where there is no uncertainty about distance. In this type of measurement, the observer judges whether the distance separating a pair of lines or points is greater or smaller than the standard separation – a simple size judgment for targets presented at a fixed viewing distance. The Weber fraction for lateral separation is 2–4% (Burbeck, 1987a; Klein & Levi, 1987; McKee, Welch, Taylor, & Bowne, 1990). Therefore, we might expect the Weber fraction for objective size judgments, involving comparisons at different distances, to be greater than 2–4% because of additional noise from the distance estimates. That is true only if the noise in the distance estimates is comparable to or larger than the noise in the estimates of retinal subtense. As we describe next, distance estimates are quite imprecise.

### *Estimating distance*

Every introductory psychology book lists a large number of cues to distance. These include binocular parallax, accommodation, motion parallax, interposition (occlusion), familiar size cues, aerial and linear perspective, and texture gradients.<sup>2</sup> We will not attempt to describe in detail how each is used to estimate depth or how they are used in combination (see Chapter 15, this volume; see also Landy, Maloney, Johnston, & Young, 1995; Yuille & Bülthoff, 1993). Instead, we are concerned with the range, accuracy, and precision of each cue because these limitations affect size constancy. Certainly, not all cues are equally useful in estimating distance. Interposition specifies only which object is in front of the other, not the distances separating them. Familiar size relies on exact knowledge of the dimensions of a recognized object to provide a distance scale. Accommodation works only at short distances because it is driven by the defocus produced by objects lying outside the focal plane. Human sensitivity for defocus is roughly 0.2–0.4 diopter under optimum conditions (Campbell, 1957; Legge, Mullen, Woo, & Campbell, 1987), so accommodative information about distance is constrained to 2 m or less.<sup>3</sup>

Binocular parallax,<sup>4</sup> sometimes called *convergence angle*, is potentially the most powerful cue to object distance. Information about binocular parallax can be obtained either indirectly, from the neural signal transmitted to the convergence system to minimize disparity, or directly from the sensed position of the eyes. Foley (1980) estimated the precision of parallax judgments as about 5 arcmin at 4 m, or roughly 10% of the parallax angle. At larger distances, 5 arcmin of disparity translates into a very large uncertainty about linear distance, so binocular parallax is not very useful at distances much beyond 8 m. Distance information derived from binocular parallax is known to be *inaccurate*. When based on parallax alone, the perceived distance of a near object exceeds its physical distance, whereas the distance of a far object is seen as less than its physical distance (Foley, 1980).

Relative disparity, that is, the difference in the disparity of two features, can be used to judge the distance separating objects. However, the interpretation of a given disparity difference depends on the estimated viewing distance; 10 min of disparity at a viewing distance of 1 m corresponds to a much smaller objective distance than 10 min of disparity at a viewing distance of 5 m. Inaccuracies in estimating distance from binocular parallax will therefore affect the accuracy of relative disparity judgments as well (Foley, 1980; Norman, Todd, Perotti, & Tittle, 1996; chapter 15, this volume). As one example, Johnston (1991) reported that a three-dimensional shape defined only by disparity appeared thicker or thinner (its dimension along the z-axis) at different viewing distances. Relative disparity judgments, even those made at a fixed viewing distance, are also not very *precise*. Under the best circumstances, the Weber fraction for relative dis-

parity is 5–6%, and it is often found to be much higher (McKee, Levi, & Bowne, 1990; Norman et al., 1996).

When the head translates, objects at different distances move at different speeds on the retina – motion parallax. If the motion is self-generated, so that the observer has some way of calibrating relative speed, motion parallax can be a robust cue to distance. Available evidence indicates that, by itself, motion parallax produces fairly accurate estimates of relative distance (Huber & Davies, 1995; Landy et al., 1995; Rogers & Collett, 1989; Rogers & Graham, 1979). The precision of distance information derived from motion parallax depends on speed discrimination. The Weber fraction for moderately fast angular speeds (>3 deg/sec) is 5–8%. Because normal head movements are not very fast, distant objects may move at speeds considerably slower than 3 deg/sec where the Weber fraction for speeds is much higher, meaning that estimates of distance based on motion parallax would be less precise.

The texture elements that define a surface gradually decrease in angular subtense with increasing distance. As noted earlier, this texture gradient can be used to judge relative distance. Linear perspective depends on a similar decrease in the angle between straight contours extending toward the horizon. Given a scaling factor that specifies the distance associated with a particular angular subtense, the observer could, in principle, judge the physical distance to any object sitting on a flat, regularly textured surface. Unfortunately, changes in angular subtense produced by surface tilt or by large-scale physical irregularities, such as a change from pebbles to boulders, are necessarily confounded with the angular changes produced by distance. These confounds, as well as the possible inaccuracy of the scaling factor, compromise the accuracy of texture cues to distance in natural environments. Texture and perspective cues do, however, supply the most precise information about distance because they are based on the same precise information used for angular judgments of lateral separation. As noted earlier, the precision of angular judgments is roughly 2–4%.

In reduced viewing conditions, inaccuracies in estimating distance should necessarily lead to systematic biases in size constancy. Such biases may be less likely in natural environments where distance information from one source could be used to correct distance information from another source (Landy et al., 1995). Improving the precision of distance estimates is more problematic because there is no way to reduce the inherent noise associated with these estimates. In circumstances where many cues to depth are available (short distances and natural surroundings), probability summation among independent estimates of distance could improve precision. However, experimental measurements show that depth judgments in multi-cue conditions are not generally more precise than single-cue depth judgments (Norman et al., 1996).

If we assume that the estimate of the angular subtense (the proximal stimulus)

and the distance estimate are independent, we can use a simple propagation of error approach<sup>5</sup> to predict the precision of size constancy. For example, the predicted precision of size constancy is about 5.4% based on the most precise Weber fractions for angular subtense (2%) and motion parallax (5%). Using Foley's estimate of the precision of binocular parallax (10%) produces a corresponding increase in the predicted threshold to slightly over 10%. On the other hand, if the observer can utilize texture information to estimate distance (2%), size constancy could achieve a precision of less than 3%. The precision of size constancy judgments can thus reveal much about how distance information is integrated into the size estimate.

#### *The accuracy and precision of objective size judgments*

The first experimental measurements of size constancy were made in the late nineteenth century by Martius in Wundt's laboratory (Boring, 1942). Martius reported perfect size constancy, but subsequent measurements by Thouless (1931) indicated that perceived size was a compromise between angular subtense and physical size. In an excellent and often cited study, Holway and Boring (1941) examined how various sources of depth information influenced perceived size. Five observers made repeated matches of a circular test target presented at distances ranging from 10 to 120 ft. The test, actually a circular patch of light projected on a screen, subtended 1 degree at all viewing distances. The observer sat at the intersection of two dimly lit corridors and adjusted the comparison target, 10 ft away in one corridor, until it matched the test target in the other corridor (graduate students take note: all measurements were made after midnight!). Available depth information was successively reduced from a "full-cue" binocular condition to a condition that employed monocular viewing, then monocular viewing through an artificial pupil, and finally monocular viewing, an artificial pupil, and a long reduction tunnel made of heavy black cloth. The use of an artificial pupil (1.8 mm) to reduce distance cues is rather curious because, although it would minimize accommodation cues, accommodation would not be much use for the large distances employed in this study. More likely, the artificial pupil made it difficult for the observers to estimate relative size and texture cues in the low illumination, the cues that the reduction tunnel was expected to eliminate.

As Figure 14.2 shows, size constancy was accurate when adequate information about depth was available. Predictably at these large viewing distances, binocular parallax did not confer any benefit over good monocular depth information. In fact, when the test target was viewed binocularly, its objective size was slightly overestimated, although this pattern was not universal (note the differences between the observers in the small graphs at the bottom of Figure 14.2). Some observers, relying on binocular parallax at only the short distances

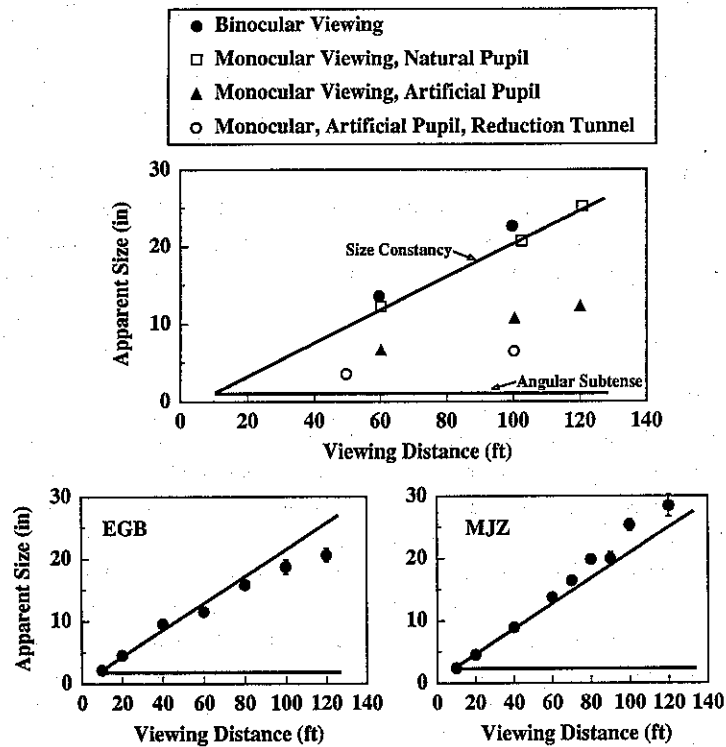


Figure 14.2. Results from the classic size constancy study by Holway and Boring (1941). Observers matched the size of a comparison disc viewed at a fixed distance to the objective size of a test disc viewed at various distances. The upper graph shows pooled results for the various experimental conditions listed in the top box. The lower graphs show individual results from two observers in the binocular condition.

where it is useful, may have underestimated the 10-ft distance to the comparison target and so systematically increased the size of their matches. Contrary to the expectations of the experimenters, observers never adjusted the comparison target to match the angular subtense of the test target, even in the most reduced condition. Fragmentary information about linear perspective and texture from the dimly illuminated reduction tunnel was sufficient to promote some tendency toward constancy. Subsequent studies managed to eliminate the residual light reflected from the surroundings and obtained perfect angular matches (Lichten & Lurie, 1950; Over, 1960).

Judgments of objective size at great distances are surprisingly accurate. Gibson (1950) asked observers to choose which of a set of nearby wooden posts matched a similar barely visible post located more than 2,000 ft away. The judgments were made in a barren open field and, at 2,000 ft, the test post

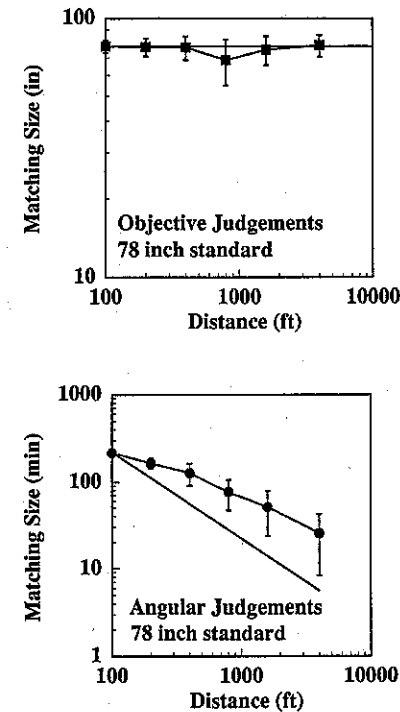


Figure 14.3. Results from the Gilinsky (1955) study showing observers' ability to match either the objective size of a test target (upper graph) or the angular subtense of a test object, depending on instructions from the experimenter. Data pooled from all observers.

subtended only about 8 arc minutes. Despite the difficulty of the judgment, observers chose the correct post. Gilinsky (1955) extended the range to an even greater distance (4,000 ft) and again found remarkable accuracy. In addition, Gilinsky asked her observers to match the angles subtended by the distant test objects; these matches were much less accurate than the objective matches (see Figure 14.3). Generally, her observers overestimated angular subtense, a result confirmed in a later study by Leibowitz and Harvey (1967).

The constancy judgments in the Holway-Boring study were also quite *precise*. In Figure 14.4, we have plotted "pseudo-Weber" fractions (mean variation/mean) for the two observers who participated in all experimental conditions. These values are averaged over all test distances; individual fractions for particular distances were even better. In fact, the best pseudo-Weber fractions for size constancy were about 2%, equal to the best Weber fractions for lateral separation measured in a single plane. In a later study, Burbeck (1987b) confirmed this Holway-Boring result. Using a standard increment threshold para-

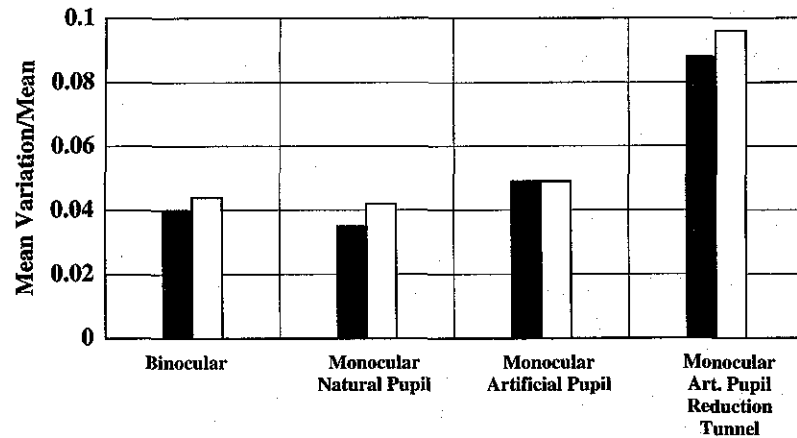


Figure 14.4. Precision measures for size constancy derived from data presented in Holway and Boring (1941). Mean variation is divided by the mean stimulus value for the two observers (represented by black and white bars) who participated in all conditions.

digm and the contemporary stimulus of choice, a sinusoidal grating, Burbeck measured spatial frequency discrimination for gratings presented at the same distance on two adjacent cathode ray tube (CRT) screens. The spatial frequency in cycles per degree of the gratings differed only by a small amount – the tested increment. Guided by feedback, Burbeck's observers judged which grating had the higher spatial frequency. She then moved one of the screens to twice the distance of the other and repeated the threshold measurements. Although the basic spatial frequency of the gratings in angular units (cycles/degree) now differed by a factor of 2, the thresholds were nearly identical to those measured when the screens were at the same distance. In short, the Weber fraction for objective spatial frequency (cycles/cm) was about 3%, comparable to the best judgments of lateral separation made in the fixation plane.

Burbeck also asked her observers to judge small differences in the angular spatial frequency (cyc/deg) of the targets presented at two different distances. Initially, observers found the angular judgments almost impossible, but with feedback and considerable practice, they were able to perform with a precision about equal to that of their objective judgments. Burbeck argued that the angular frequency was indirectly estimated from objective frequency information. She concluded that observers have no direct access to spatial frequency, or to any other kind of spatial information, expressed in angular units. This provocative conclusion is consistent with the Gilinsky results shown in Figure 14.3, but it is difficult to reconcile with the imprecision of distance judgments. If the observer has access only to size information that is automatically scaled for viewing distance, why are size judgments so precise when distance judg-

### Size Constancy from Relative Size

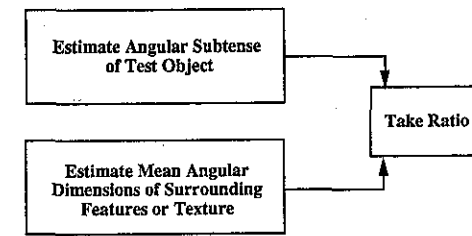


Figure 14.5. Flow chart showing the operations used to estimate objective size from angular relationships.

ments are so much noisier? A texture cue might account for the precision of Burbeck's results. Do her results imply that all size information is automatically scaled by depth calculated from texture gradients? Even if the texture-gradient cue to depth accounts for the precision of size constancy, it is unlikely that texture gradients can provide depth information of sufficient accuracy to explain the accuracy of size constancy at great distances, for example, the Gilinsky results.

The accuracy and precision of size constancy judgments force a different conclusion. In natural circumstances, observers do not use depth estimates to calculate objective size. Instead, they rely on the relational determination of size, a cue that can supply accurate, precise information about relative size without requiring an independent estimate of distance (Gibson, 1950; Hochberg, 1972; Rock, 1977; Rock & Ebenholtz, 1959; Sedgwick, 1986).

#### Relative size

The mean angular dimensions of our surroundings decrease with increasing distance, so the ratio of the angle subtended by an object to the mean of the angles subtended by its surroundings is fairly invariant with viewing distance. The observer can equate the objective size of two objects, as required in size constancy experiments, by simply equating the ratio of the test and comparison objects to their respective surroundings (Figure 14.5). When the surroundings contain a regularly textured surface, such as a floor covered with tiles, the texture supplies a local metric that is automatically scaled with distance; the dimension of an object sitting on the textured surface can be measured in the number of contiguous texture elements and thus equated to other objects at other positions on the surface (Gibson, 1950; Nakayama, 1994). Note that, in this case, texture

is *not* being used to estimate distance; instead, it is serving as a local ruler for size estimates.

Relational effects have a measurable influence on size judgments, even with binocular viewing at small distances. Rock and Ebenholtz (1959) asked observers to match the length of a line surrounded by a small rectangle to a similar line surrounded by a rectangle three times larger. The observer had to turn 180 degrees to view each of the rectangular configurations in succession; the two configurations were self-luminous and were presented in total darkness at the same fixed distance of 5 ft. Although a few observers matched the physical size of the lines, about half of the group made near-perfect relational matches, that is, the line in the smaller rectangle was shortened so that it was proportionally the same length as the line in the larger rectangle. Is this result due to individual differences in the interpretation of the experimental instructions? Maybe. When Wenderoth (1976) replicated the Rock–Ebenholtz study with instructions that stressed a match based on physical equality, none of the observers made perfect relational matches. Nevertheless, despite instructions to the contrary, Wenderoth's observers still reported that the lines were equal when the line in the smaller rectangle was, on average, 15% shorter than a true physical match. In both studies, observers recognized that the two configurations were at the same distance, but their size judgments were influenced by relational effects nonetheless.

The effect found in Wenderoth's study is about the same magnitude as the size misjudgments found in many geometric illusions. For example, in the Ponzo illusion (Figure 14.6A), the upper line looks longer than the lower line, although the two lines are actually the same physical length. Gillam (1973) found that observers equated the two line lengths in the Ponzo illusion when the physical length of the lower line was about 14% more than that of the upper line, a value very close to Wenderoth's result. Gregory (1963, 1973) argued that geometric illusions of this type are examples of inappropriate constancy scaling. Although observers know that the two short horizontal lines are on a frontoparallel surface at the same distance, the linear perspective cues associated with the two converging lines trigger an automatic rescaling of size.

Humphrey and Morgan (1965) challenged Gregory's idea by inventing a clever variant of the Ponzo illusion in which the horizontal lines were simply rotated 90 degrees (Figure 14.6B). The two vertical lines now appeared to be the same length, despite the presence of the perspective cues. If the observer had calculated depth from the perspective cues and then rescaled all other features, the vertical lines should have been affected in the same way as the horizontal lines. However, as noted earlier, relational size effects do not require the observer to estimate or calculate depth. Instead, the effects depend only on the ratio of local angular measurements – in this case, the ratio of the horizontal line length to the local separation between the converging lines. As Gillam

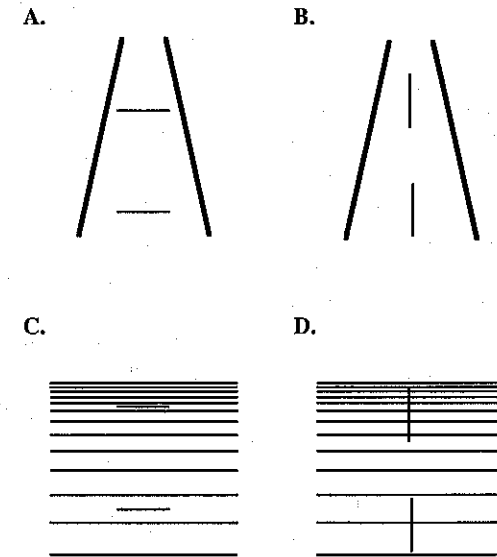


Figure 14.6. (A) The Ponzo illusion; horizontal bars are the same length. (B) Humphrey–Morgan variant, with vertical bars of the same length that appear the same length. (C,D) Illustrations based on Gillam (1973) showing that foreshortening affects the apparent length of vertical bars (D) but not of horizontal bars.

(1973) has reasoned, in perspective processing the lateral distances between the converging lines represent equal horizontal distances, so in the Ponzo illusion only the size of horizontal dimensions should be affected. On the other hand, as Gillam has also shown, texture foreshortening of the type shown in the lower half of Figure 14.6 differentially affects vertical dimensions because the observer interprets the spaces between each pair of long horizontal lines as representing an equal vertical distance. In this case, the illusion works for the orthogonal direction; the upper *vertical* line appears longer than the lower *vertical* line, although once again, the vertical lines are of equivalent physical length. When the lines are rotated 90 degrees, the illusion disappears (compare Figures 14.6C and 14.6D).

Geometric illusions may be based on processes that are unrelated to size constancy. Still, several observations make Gregory's hypothesis at least plausible. First, if an elaborate pictorial representation of depth is added around the converging lines in the Ponzo illusion, the perceived illusion is even greater (Coren & Girgus, 1978; Sedgwick & Nicholls, 1993). Second, the accuracy and precision of size constancy at long distances are most easily explained by assuming that the observer uses relational cues, and because this relational process is largely automatic, it could be misapplied. Finally, Gillam (1973) noted that

foreshortening has a different dependence on distance than does linear perspective. She found that the magnitude of the illusion in the foreshortening configuration (Figure 14.6D) was correspondingly smaller than that found in the perspective configuration (Figure 14.6C), as predicted from the differential effects of these two pictorial cues. Gillam's results suggest that visual processing of relative size is remarkably subtle and is tightly bound to an implicit understanding of how texture and perspective change with distance. So, although relational effects do not depend on a distance calculation, their power in affecting perceived size probably derives from their association with distance.

The precision of size constancy in the Burbeck (1987b) study (Weber fractions of 2–4%) suggests that her observers were basing their judgments on the relative size cue – the relative width of the grating bars to the angle subtended by the monitor screen – a cue that was not available for the angular judgments in her study. Still, this conclusion raises a puzzling question: Why were her observers able to use angular relationships, that is, the angular ratio of bars to screen, to judge objective size, but found it difficult to compare the angular spatial frequency of the sinusoid on one screen to that on the other screen? Presumably, the relational effects must interfere with direct encoding of angular subtense (as in the Ponzo illusion), with the result that judgments of angular subtense in natural contexts are often inaccurate. This interference is not inescapable because, with feedback, observers can learn to make precise judgments of angular size despite their percepts.

Does the observer rely exclusively on relational size information when judging the objective size of visual features? Obviously not. By itself, relative size produces only small changes in perceived size. Consistency with other depth information is required to generate true size constancy, that is, the percept that two identical objects at different distances are the same size. Consider the Ponzo illusion once again. The horizontal lines in Figure 14.7A do not appear equally long, but neither do the lines in Figure 14.7B, which are matched in terms of relative size (same proportion of the lateral separation between the tilted lines). In this situation, where all the other depth information asserts that the pictured lines are on a frontoparallel surface, the relational effects are minimal. In Figure 14.7C, the lower line has been increased in length by only 15% and now looks about equal to the upper line. Even if the converging lines in Figure 14.7B were the baseboards of a real hallway, and the horizontal lines were identical decorative markings on widely separated floor tiles, they might not *appear* exactly identical. Still, the depth information would likely induce a percept of near equality. If an observer were asked to draw a matching line, equal to one down the hall, the relational information supplied by the converging diagonals would improve both the accuracy and the precision of the match. You can check your own ability to judge relative size. Is the lower line of Figure 14.7D the same

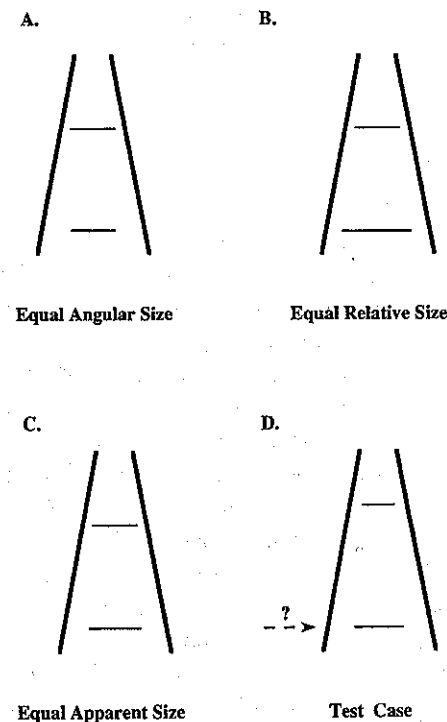


Figure 14.7. (A) Ponzo illusion with bars of the same angular subtense. (B) Horizontal bars are of the same proportional length in relation to the lateral separation between tilted lines, yet they do not appear equally long. (C) The lower horizontal bar is 15% longer than the upper bar and now appears to be the same length. (D) Do the horizontal bars have the same proportional length (as in B)?

relative length as the upper line? If not, is it too large or too small? Try drawing a line in the correct relationship if you think the lower line is in error.<sup>6</sup> Now judge the relative depth of these two horizontal lines using the information from the linear perspective cues. You will probably find that the depth judgment is more difficult than the relative size judgment. Using texture gradients to judge relative distance, that is, whether one object is three times as far away as another, is undoubtedly more difficult than judging the equivalence of local ratios for each object. Therefore, size constancy can be extremely precise under circumstances where the depth information is too noisy or inaccurate to supply a scaling factor of the requisite precision. Two predictions follow from this argument. First, an observer's depth and size estimates will not always be strongly correlated – a result often observed in studies of size constancy (see chapter 18, this volume; Epstein, Park, & Casey, 1961; Sedgwick, 1986). Second, if the texture



or frame surrounding the test object and the comparison object are quite different, observers should make systematic errors, an effect evident in the study by Norman et al. (1995).

#### *Size constancy from distance information*

In natural full-cue conditions, distance information appears to play a "supporting role" in size constancy. Size judgments are based on the ratio of the angles subtended by objects with respect to their surroundings, and distance information supplies a kind of consistency check. However, in the absence of adequate relational information, distance cues are sufficient to promote size constancy.

Almost any isolated cue to distance is somewhat effective in laboratory settings. When accommodation and convergence supplied the only depth information, Harvey and Leibowitz (1967) found that observers were moderately accurate in matching the objective size of rods viewed at a distance of less than 1 m. Hell (1978) found that monocularly viewed rods, presented with their ends obscured in an otherwise empty visual field at different distances, were matched on the basis of their angular subtense, provided that the observer's head was stationary. However, when the head was moved laterally, the matches fell halfway between angular and objective sizes, showing that motion parallax can produce some degree of size constancy. Binocular viewing is unnecessary for constancy in natural settings, but relative disparity alone can produce fairly accurate and precise estimates of objective size, provided that the viewing distance is fixed (McKee & Welch, 1992).

Other minor cues can also influence size constancy. For example, Gregory, Wallace, and Campbell (1959) showed that knowledge of our own movements affects perceived size. The apparent size of an afterimage is usually determined by its angular subtense relative to the apparent distance of the surface on which it is "projected." When projected on a nearby surface, it appears much smaller than when projected on a more distant one, despite its fixed angular subtense. This effect is known as *Emmert's law* (Gregory, 1987). Gregory et al. (1959) noticed that afterimages viewed in total darkness appeared to change size when the head was moved forward and back. Thus, information about changes in distance associated with voluntary movements can be used to scale size, at least when no other information about distance or relative size is available. Knowledge about the size of a familiar object affects judgments when the object is viewed monocularly in total darkness. For example, enlarged versions of common coins are underestimated in these circumstances (Epstein, 1967).

We argued earlier that relational effects interfered with the accuracy and precision of angular size judgments. What happens to angular judgments in the absence of any information about relative size? In their study of size constancy based on disparity alone, McKee and Welch (1992) compared objective and

angular size thresholds. Observers were asked to discriminate small changes in the vertical distance separating a pair of horizontal lines. From trial to trial, the target was presented at random at one of nine disparities spanning a range of  $\pm 40$  arcmin centered on the fixation plane. The lines were actually displayed at a fixed distance in a stereoscope, but the changes in disparity created compelling changes in apparent depth. For the objective size judgments, the vertical separation between the lines in angular units was scaled, based on calculations from relative disparity, as though the physical distance to the target were really changing. Guided by error feedback, observers were required to judge the vertical separation in physical units (height in centimeters). For the angular judgments, the mean angular subtense was fixed with changes in disparity, and observers used feedback to judge incremental changes in the angle separating the lines. As shown in Figure 14.8, angular thresholds were somewhat better than objective thresholds, although this difference was not significant at large separations. This result argues that angular size is *not* calculated indirectly from objective size.

Discrimination thresholds for targets presented only in the fixation plane (zero disparity) were also included in this study. These fixation plane thresholds were consistently better than the angular thresholds, providing additional evidence that observers do not have access to the "angle on the retina," or else the angular and fixation plane thresholds would be identical. Individuals with normal stereopsis only have a representation of size or location that is mediated by their binocular system – by the fusion of the signals from both eyes' retinæ. The data in Figure 14.8 were obtained for a duration too brief to permit a change in convergence (150 msec). Therefore, separation judgments for targets presented off the fixation plane were mediated by disparity mechanisms responsive to nonzero disparities. These mechanisms are less sensitive to size than mechanisms that respond only to targets in the fixation plane, accounting for part of the loss in the precision of the angular judgments. However, even when the targets were presented for a longer duration (1500 msec) and observers were encouraged to converge to the targets presented at different depths, angular size thresholds remained slightly higher than thresholds for targets presented only in a single plane. McKee and Welch speculated that interference from size constancy scaling might have produced the small decrement in performance.

The surprising result is that the objective size thresholds were so precise (~6.5%). Disparity increment thresholds are known to be 10% or greater for durations as short as 150 msec (McKee, Levi, & Bowne, 1990). Thus, if objective size were calculated by combining the disparity estimate with the angular size estimate, the thresholds should be much higher. The explanation may be that the observers were not making an exact estimate of disparity. Instead, they could have used the widely separated depth planes as a code to rescale the angular estimate; for example, the second plane from the front requires the

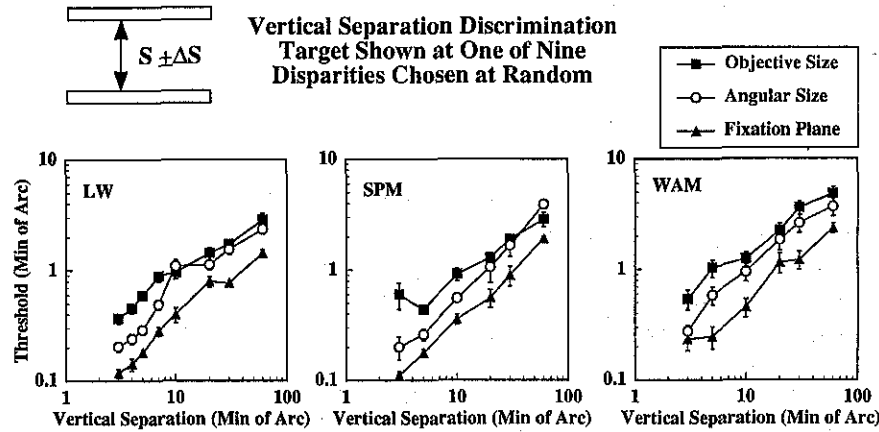


Figure 14.8. Data from McKee and Welch (1992). Increment thresholds for judging vertical separation between lines diagrammed at the top of figure for three observers. Squares show objective size thresholds measured when angular separation was scaled by disparity as though the physical distance to the target was changing; targets were presented at random at one of nine different disparities from trial to trial (see text). Open circles show angular size thresholds for unscaled targets presented in the same random disparity condition. Triangles show size thresholds for targets presented in the fixation plane only.

second largest scale. With feedback, observers are very good at using multiple implicit “standards.” Morgan (1992) asked observers to judge relative width for targets presented randomly at different orientations from trial to trial. Each orientation had a different implicit standard: for example, for vertical,  $10 \text{ arcmin} \pm \Delta$ ; for oblique,  $11 \text{ arcmin} \pm \Delta$ ; for horizontal,  $12 \text{ arcmin} \pm \Delta$ . The observers were never shown the standard width for any of the orientations; instead, they learned a different internalized standard for each orientation. These multiple-standard width judgments were nearly as good as judgments made with a single standard width. In an earlier study on velocity constancy, McKee and Welch (1989) asked observers to label 10 widely separated depth planes with a number from 0 to 9. After a small amount of practice, observers accurately labeled each depth plane on about one-third of the trials, and they were seldom off by more than one adjacent plane when they mislabeled the planes. This pattern of errors should increase the single-plane threshold by about a factor of 1.5, according to calculations. In the McKee–Welch study of size constancy, the observed increase from the single-plane threshold ( $\sim 3.5\%$ ) to the objective-size threshold ( $\sim 6.5\%$ ) is close to this prediction. Thus, it is certainly possible that the precision of the objective size thresholds was based on a learned code for rescaling.

Is this result merely a laboratory curiosity? Perhaps, but in familiar settings

(office, laboratory, kitchen, playing field), a learned code for rescaling might prove useful, given the imprecision or inaccuracy of extant depth estimates. We next consider an unusual example of recoding size judgments.

#### Learning constancy

Although we have a number of “hard-wired” neural systems for estimating relative depth (disparity detectors, motion detectors, and the like), size constancy is undoubtedly based on our long experience with real objects and surfaces – on a learned calibration. Depth judgments are affected by ongoing changes in the reliability of different sources of information (Young, Landy, & Maloney, 1993), so active recalibration of size may also occur. Can observers learn to make accurate size judgments based on orderly but *unnatural* information?

In the McKee–Welch study described earlier, separate psychometric functions were generated for each of the nine tested disparities; the point of subjective equality (the stimulus value corresponding to the 50th percentile on each psychometric function) was taken as a measure of perceived size. As shown by the open symbols in the upper graphs in Figure 14.9, the two observers were reasonably accurate in judging objective size. McKee and Welch next exactly inverted the natural relationship so that the angular separation *decreased* as the target appeared closer in depth and *increased* as the target appeared farther away; the angular separation for  $+40 \text{ arcmin}$  of disparity in their “anti-constancy” experiment was set equal to the separation used for  $-40 \text{ arcmin}$  in their constancy experiment. In brief experimental sessions taken over a couple of days (600 trials total), observers practiced judging objective size, using error feedback to recalibrate their size judgments in this unnatural condition. Then they made the measurements shown in Figure 14.9 (filled symbols in upper graphs). Surprisingly, the anti-constancy condition produced results that were just as accurate as those in the constancy condition. The anti-constancy judgments were less precise (lower graph in Figure 14.9), but, with more than 2 days of practice, the errors might have reached the same level as the constancy measurements.

Did the target seem to be the same size in the anti-constancy condition? No. The inverted relationship was, in fact, exaggerated by natural size constancy, so that the distant target looked enormous compared to the puny distance separating the lines at the nearest disparity. Even weeks of practice could not have overcome continuous natural feedback about the real relationship between relative disparity and angular subtense. Nevertheless, these results suggest that there is a flexible calibration mechanism involved in size constancy. If observers had been immersed in a virtual reality domain where the natural relationship between depth and angular subtense was universally inverted, the recalibration might have become sufficiently automatic to foster an *appearance* of constant size.

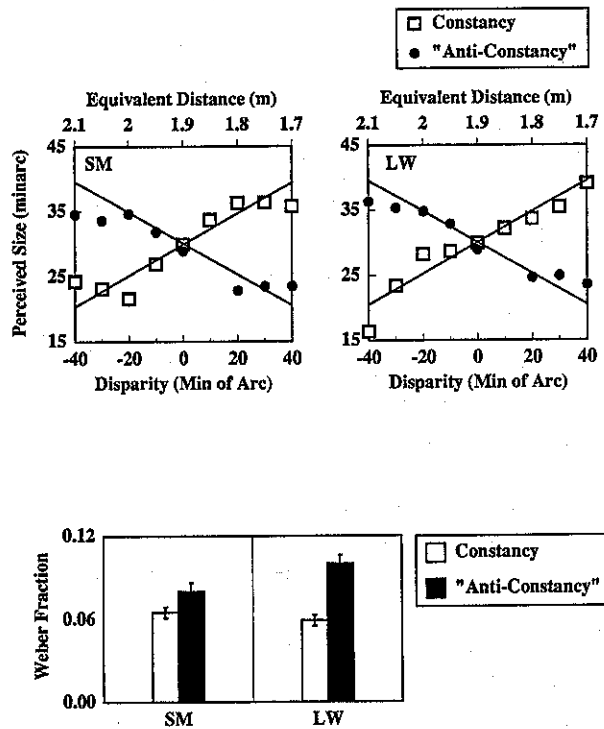


Figure 14.9. Data from McKee and Welch (1992). Upper graphs show PSEs for two subjects for the constancy and anti-constancy conditions. Open squares are based on the condition in which angular separation was scaled by disparity as though the physical distance to the target was changing; filled circles are based on the condition in which the constancy relationship was inverted so that angular subtense increased with increasing distance (see text). The lower graphs show precision measures (Weber fractions) for the same conditions. Target separation in the fixation plane was 30 arcmin.

Indeed, work from the Ross laboratory on the size adaptation experienced by scuba divers indicates that as little as 20 min of altered underwater optics can affect apparent size in normal viewing conditions for a few minutes after leaving the water (Franklin, Ross, & Weltman, 1970; Ross, Franklin, Weltman, & Lennie, 1970).

*Dual calculations?*

Certain kinds of information are lost in the course of visual processing. We can discriminate between wavelengths, but we have no knowledge of the separate signals in the three types of cones. We can see changes in disparity, but we cannot identify the contribution of each eye to the fused image (Templeton &

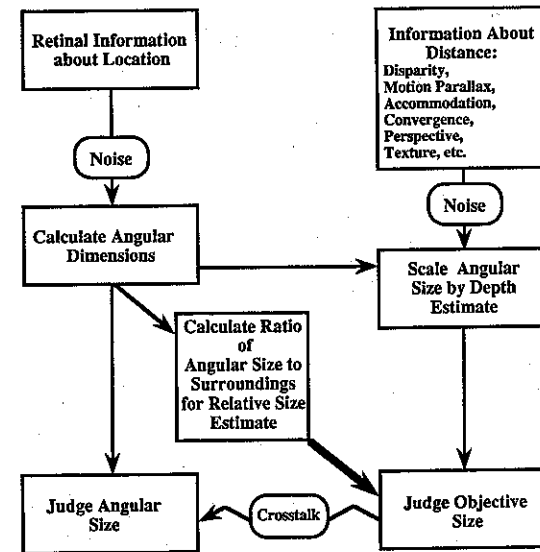


Figure 14.10. Flow chart showing dual calculations of angular and objective size, as described in text.

Green, 1968). Is size constancy like that? Have we lost information about angular subtense in the process of constructing a representation of true size? The results from the McKee-Welch study indicate that "angle on the retina" is not a given. Like all other information about the physical world, it must be translated from the light distribution on the retina into a neural representation of relative positions and lateral distances. This neural calculation could automatically include the scaling required for size constancy, such that subsequent stages would have no access to angular information. Apparently, it does not. There are far too many indications that we also have access to good information about angular size.

Rather than the simple dichotomy diagrammed in Figure 14.1, the human brain must be simultaneously calculating both the angular and objective dimensions of the whole visual scene (Rock, 1977). The flow chart in Figure 14.10 summarizes our view of these dual calculations. Size constancy depends on two separate processes: (1) angular size scaled by depth and (2) relative size calculated from the ratio of local angular measurements (Hochberg, 1972). The relative size calculation provides the most precise information, but in the absence of depth signals from the other processing stream, relative size is not sufficient to promote accurate judgments of objective size. As one would expect, size constancy is promoted or enhanced by the concurrence of many cues to depth.

The noise sources, indicated by the oval shapes in Figure 14.10, are primarily

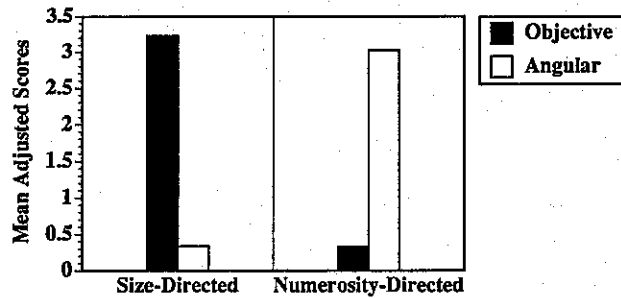
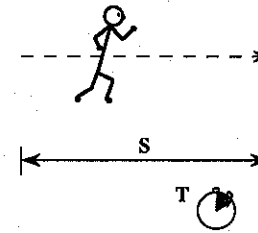


Figure 14.11. Data from Epstein and Broota (1986) showing the effect of attention on whether observers make objective size matches (dark bars) or angular size matches. When attention is directed to the size of the square, they make objective size matches. When attention is directed to the number of dots on the square, they make angular size matches of the square.

associated with the low-level detectors for retinal location (local sign) and distance (disparity, motion parallax, convergence, accommodation, texture coding). We suspect that the neural calculation that translates angular dimensions into objective dimensions is very efficient. For example, in Figure 14.8, objective thresholds are slightly higher than angular thresholds, a difference that is exaggerated at small separations; this pattern is consistent with an additive source of noise. The calculation of objective size from angular dimensions at various disparities adds about 20 arcsec of uncertainty to the objective thresholds. We also concede that the automatic scaling of size with distance does interfere slightly with the calculation of angular size in natural surroundings, a noise source we have labeled "crosstalk" in the flow chart.

What determines which of the dual calculations dominates size perception? To some extent, it depends on where the observer's attention is directed. Epstein and Broota (1986) used different tasks to direct attention either to the size of an object or to the markings on it. They presented observers with posterboard squares of various sizes, each covered with a variable array of randomly positioned dots. In one condition, observers were asked to judge the size of the squares presented briefly at various distances. After each square had disappeared, observers marked a test sheet that pictured potential choices for a match; the choices included the objective size of the square, the angular subtense, and some intermediary sizes. In the other condition, the observers were asked to judge, as quickly as possible, whether the number of dots on the square was odd or even. After they had made several numerosity judgments, they were asked to judge the size of the square seen on the last trial. Epstein and Broota compared the matches chosen in the last trial for the two conditions. When attention was directed to the size of the square, observers matched the square on the basis of objective size, but when their attention was directed to the dots, they chose a

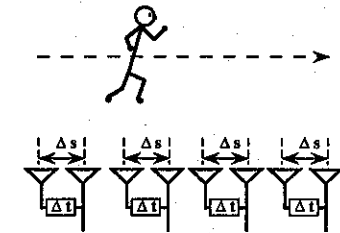
#### Indirect Calculation of Angular Speed from Distance and Time Estimates



$$V = \frac{\text{Distance Traversed (S)}}{\text{Time Taken for Traverse (T)}}$$

Use size constancy scaling to convert distance traversed from angular units (deg) into objective units (m)

#### Direct Calculation of Angular Speed from Signals in Motion Detectors



$$V = \Delta s / \Delta t$$

Local motion detectors encode the time taken to cross the distance separating pair of receivers. Angular speed (deg/sec) could be converted directly into objective speed (m/sec).

Figure 14.12. Two procedures for calculating speed. In the left half of the figure, speed is calculated from the distance traversed by the target and the time taken for traverse. In the right half of the figure, speed is calculated from motion detectors that perform the calculation for distances smaller than the whole traverse (see text). Speed constancy could be based on rescaling information from either procedure.

match on the basis of angular subtense (see Figure 14.11). This result makes sense if we consider that angular information is used to specify the "background" of our surroundings, that is, to specify the texture and perspective cues that provide information about relative depth. An object of interest is the "figure" on this "background," and generally, we want to know its true physical size in order to decide the appropriate behavioral response. Is the beast before us a cat or a tiger? Thus, these parallel calculations serve different functions in guiding human action, and apparently, we need some awareness of both for our own well-being.

### Speed constancy

As you walk away from a moving object, the retinal speed of the object decreases, but it does not appear to slow down. *Speed constancy* refers to the human ability to compensate for the changes in angular velocity associated with changes in viewing distance and thereby to maintain an invariant estimate of objective speed. It is usually treated as an extension of size constancy because of an assumption about how angular speed is encoded by the human visual system. As shown on the left side of Figure 14.12, speed could be calculated from separate measurements of the distance traversed by the moving target and

of the time taken to cover that distance. If the distance measurement were scaled by the same neural calculation used for size constancy, perceived speed would be invariant with viewing distance. In fact, this formulation of speed constancy is nearly correct. What is incorrect is the scheme for encoding angular speed.

When you take a motor trip, you don't have to wait until you've completed your journey to know how fast your car was moving. Instead, you read the car's speedometer. Similarly, the motion system can estimate speed from the signals generated while the target is in transit. There is abundant physiological and psychophysical evidence for specialized motion mechanisms that measure space and time on a small scale (see the right side of Figure 14.12). There are also psychophysical data that argue explicitly against an indirect estimate of speed from traverse length and target duration. McKee (1981) measured speed discrimination for targets that moved across a fixed aperture. Because the length of the traverse was fixed, the time the target spent crossing the aperture necessarily varied with target speed, so observers could have judged speed on the basis of duration. However, McKee showed that the speed judgments for the moving targets were more precise than comparable duration judgments for static targets (a result confirmed by Orban, de Wolf, & Maes, 1984). In short, speed discrimination is based on a temporal signal that is more precise than the time estimate associated with the whole traverse.

In contrast to these findings, Mandriota, Mintz, and Notterman (1962) had earlier shown that random variations in the distance traversed by the target elevated speed discrimination thresholds dramatically. McKee noted similar difficulties initially when she randomized the length of the traverse from trial to trial, but with considerable practice and feedback, her observers learned to ignore these random variations and to respond as precisely as when the target traverse was fixed. Apparently, when naive observers are asked to judge how fast something is moving, they tend to pay attention to traverse length and other spatial attributes of the target; they estimate speed indirectly from lateral distance (a *static*, position-based signal) and duration (left side of Figure 14.12). We will call this speed estimate *indirect* because it is not based on *motion* signals. The indirect approach to calculating speed seems to interfere with optimal use of the motion-based signal until observers learn from feedback to respond only on the basis of this more precise signal – another case of dual calculations! We will argue that this precise motion-based signal is encoded only in angular units (deg/sec). Speed constancy, on the other hand, is based on the indirect estimate of speed.

#### *Speed constancy and spatial scaling*

In 1931, J. F. Brown published the first important study of speed constancy. As in other studies of size constancy, his observers were shown the moving test target presented at various distances and were asked to adjust the speed of a

comparison target, presented always at 1 m, until it equaled the test speed. Both targets consisted of small black squares pasted on rolls of white paper that ran between two revolving drums driven by an electric motor, all mounted within light-tight boxes. The squares were spaced such that a single square was seen at any given time moving past a fixed aperture at the front of the apparatus. Illumination from within the boxes ensured that only the moving square and its white background were visible through the apertures; otherwise, the room was darkened. Brown reported excellent speed constancy at distances ranging from 3 to 10 m and only a small deviation from an objective match (in cm/sec) at 20 m; none of the matches were based on the angular speed (deg/sec) of the test.

Because viewing was monocular in a totally dark room, it was clear that some type of relational information must account for the objective matches. Rather than assuming that speed constancy was a straightforward extension of size constancy, Brown proposed that perceived speed depended on context effects, that is, on the spatial dimensions of the moving target relative to the surroundings, particularly the framing aperture. To confirm his hypothesis, he increased all the spatial dimensions of the test display (the black squares and the aperture) by a factor of 2 and showed that the matching speed, in centimeters per second, doubled. Further increases in the spatial dimensions produced corresponding increases in matching speed. Proportionality was not perfect – scaling the dimensions by a factor of 10 increased perceived speed by only a factor of 8 – but the effects were generally consistent with Brown's hypothesis.

Brown called this phenomenon *velocity transposition* because the spatial scaling effects had induced observers to accept matches between speeds that were physically quite different, contrary to the common conception of constancy. In an essay published some 8 years later, Wallach (1939) argued forcefully that Brown had actually identified the mechanism of speed constancy. What Wallach noted was that observers accepted matches between test and comparison speeds when the ratios of the *angular* speeds to the *angles* subtended by the surrounding apertures were equal. Clearly, this condition was satisfied in Brown's constancy study because the whole test apparatus was moved to different distances, so the angular dimensions of speed and surroundings were naturally scaled together. Wallach maintained that the transposition study had triggered constancy scaling by increasing all the visible dimensions of the target, as though the test display had moved closer to the observer. Wallach found one of Brown's results particularly interesting. Brown had repeated his transposition study with *binocular* viewing and found roughly the same speed matches as for monocular viewing. Unlike relative size effects, the relative speed effects were strong enough to override contrary binocular depth information completely. Wallach concluded that, in contrast to size constancy, angular speed was not scaled by measures of depth to achieve constancy. Speed constancy depended only on relational effects – the ratio of the angular speed to the angles subtended by the surroundings.

There have been two challenges to Wallach's conception. First, Smith and Sherlock (1957) suggested that Brown's observers were actually matching the rate at which the moving dots were passing the edge of the aperture, not the speed. In contemporary jargon, they were matching the temporal frequencies of the targets. Clearly, if all spatial dimensions were increased by a factor of 2, the speed would also have to be doubled to produce the same rate past some fixed position. Smith and Sherlock (1957) demonstrated that observers could make frequency matches when the physical velocities differed considerably, a result that lent plausibility to their conjecture. Subsequent work (Diener, Wist, Dichgans, & Brandt, 1976) has shown that temporal frequency does indeed affect perceived speed. However, practiced observers can discriminate fine differences in speed despite random variations in temporal frequency (Chen, Bedell, & Frishman, 1995; McKee, Silverman, & Nakayama, 1986; Smith & Edgar, 1991); they are undoubtedly basing their judgments on the precise motion-based signal described earlier (right side of Figure 14.12) rather than the confounding dimensions that usually covary with speed changes, such as temporal frequency, distance traversed, target duration.

The second challenge came from Rock, Hill, and Fineman (1968; see also Epstein, 1978), who questioned Wallach's assertion that depth scaling played no significant role in speed constancy. With heroic experimental efforts, they managed to demonstrate some degree of speed constancy from depth alone. This study is interesting from the contrary viewpoint as well; even a hint of relational information tended to override the depth information. Indeed, subsequent work has tended to support Wallach's position. Zohary and Sittig (1993) measured speed constancy with a sparse pattern of randomly positioned dots displayed on two CRT screens. The drifting dots were viewed binocularly at two distances (1 and 2 m) but behind apertures of the same angular subtense. Observers adjusted the speed of the nearer target to match the more distant target. They easily matched the physical speeds of the dots, exhibiting good speed constancy. When Zohary and Sittig scaled the size and spacing of the dots on the near screen so that they subtended the same visual angles as the dots on the far screen, speed constancy disappeared. Observers made the matches on the basis of angular speeds despite the obvious difference in the depth of the targets. Because there was no difference in aperture size, the speed constancy found with the unscaled dots must have been based on texture scaling – on the ratio of the angular speed to the angles subtended by the moving texture. By matching the angular size of the textures, Zohary and Sittig had produced a modern variant of velocity transposition. Or was this result another case of temporal frequency matching?

The stimuli in the Zohary–Sittig study were limited-lifetime dots, appearing and then disappearing at random locations across the screen, so it was impossible

to assign a frequency rate to any one location. Nevertheless, on average, the spatial frequency<sup>7</sup> spectrum of the dots on the near screen was about half that of the unscaled dots on the far screen. In angular units, velocity (degrees per second) equals temporal frequency (cycles per second) divided by spatial frequency (cycles per degree), so temporal frequency equals velocity multiplied by spatial frequency. To match the temporal frequency of the farther dots, the angular velocity of the near target had to be doubled, exactly the compensation for distance needed to equate the physical speeds (centimeters per second) of the two displays. By scaling the size of the dots on the near screen, Zohary and Sittig had equated the spatial frequency content of the two displays; thus, the temporal frequencies were matched when the angular speeds were matched, so speed constancy disappeared. However, Zohary and Sittig also manipulated the size of the dots and their spacing separately. For some observers, speed constancy was unaffected by a change in dot spacing but diminished greatly with a change in dot size, consistent with temporal frequency coding. For others, changing either size or spacing produced matches halfway between the angular and objective speeds – less secure evidence for temporal frequency coding. One thing was clear from the Zohary–Sittig study: Depth alone was not sufficient to promote speed constancy.

McKee and Welch (1989) confirmed this result in an earlier study. They asked observers to discriminate small changes in speed while they randomly varied the target disparity. The distance traversed and the target duration were also randomly varied to encourage observers to respond on the basis of speed per se. As in their study of size constancy, they assumed that the PSEs from the psychometric functions were a measure of perceived speed. In Figure 14.13, the ratio of the PSE (angular units) to the mean speed is plotted as a function of target disparity. The oblique line shows the predicted ratios for speed constancy. In the absence of feedback or instructions, the observers spontaneously judged the speed on the basis of angular speed, not objective speed. The disparity variations had no effect on either the accuracy or the precision of their angular speed judgments.

Incorporating feedback into their experimental procedure, McKee and Welch next asked observers to discriminate small changes in objective speed (cm/sec). With feedback and practice, observers did learn to scale speed by perceived depth. However, their objective speed judgments were much less precise than angular speed judgments made with the same random variations in disparity (see the top graph in Figure 14.14). To account for the imprecision of the objective thresholds, McKee and Welch speculated that observers were estimating speed indirectly from target duration and from the objective distance (cm) traversed on every trial. They made concurrent measurements of the precision in judging the objective distance (in cm) traversed by the target–size constancy for trav-

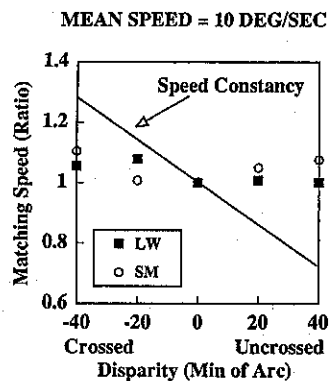


Figure 14.13. Data from McKee and Welch (1989). Ratio of matching speed to mean speed (10 deg/sec) as a function of disparity. From trial to trial, the target was presented at one of five disparities chosen at random. Observers judged whether the target was faster or slower than the mean value (method of single stimuli). No feedback was given. The tilted line shows the prediction for speed constancy scaling based on disparity.

erse length. From the precision of these measurements and the known precision of duration judgments, they were able to predict the objective speed thresholds, confirming the plausibility of their speculation. So, it is possible for human observers to use depth information for speed constancy, but they are not particularly proficient in its use.

#### Scaling angular speed or angular distance traversed?

Although Wallach (1939) stressed relational scaling of the angular speed, the calculation just described shows that relational scaling of the distance traversed would work as well. In the absence of any visible surroundings, Epstein and Cody (1980) found that the angle subtended by the traverse length alone was sufficient to promote changes in perceived speed. In this study, the test and comparison targets were single points that moved back and forth across their specified traverse distances. When the distance traversed by the test target was smaller than the distance traversed by the comparison target, observers increased the matching speed of the comparison. Epstein and Cody suggested that the distance traversed defined a relational spatial scale. Obviously, the two targets moving at the same angular speed moved the same angular distance in the same time, but the proportion of their total traverse was different. If the test target crossed the whole traverse in the time it took for the comparison target to cross half of its longer traverse, observers perceived the test to be moving faster than the comparison target, so the comparison speed was increased. Because the targets moved repetitively over the same distance ("wrap-around"), the observ-

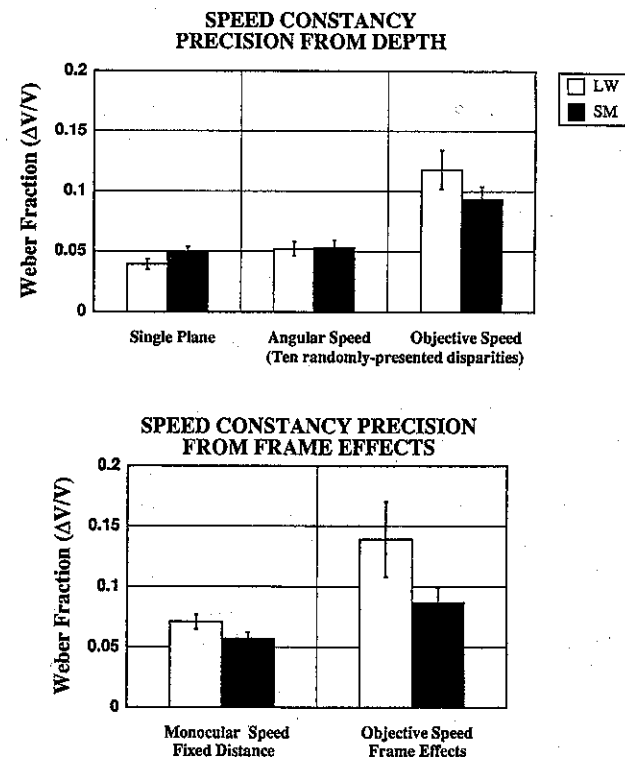


Figure 14.14. Data from McKee and Welch (1989). Upper graph: Weber fractions for speed discrimination for fixation plane targets (labeled "Single Plane"), for angular speed discrimination with random variations in disparity from trial to trial, and for objective speed discrimination with the same random variations in disparity. Feedback was given. Lower graph, left side: monocular speed discrimination averaged from two fixed viewing distances – the control condition for frame effect study. Right side: monocular speed discrimination when observers rocked back and forth from one distance to the other on alternate trials, showing frame effects on precision of speed constancy. Feedback was given.

ers might have made their matches on the basis of repetition rate. If so, they were not very good at it because the ratios of the angular speed to the angle subtended by the traverse were far from equal. When Epstein and Cody added frames scaled to the size of the traverse, the ratios were somewhat nearer to the transposition prediction but still not perfect, particularly for faster speeds. Because the frames did not change the repetition rate, perceived speed was not entirely determined by target temporal frequency. Of more importance in this context, this study shows that observers generally rely on traverse length, scaled by implicit or explicit surroundings, when asked to equate perceived speeds.

McKee and Welch (1989) also measured frame effects on speed discrimination. The observer viewed the moving target monocularly on a standard CRT

screen (the "frame") and, in the intervals between trials, rocked forward or back so that the viewing distance changed from 28 cm to 57 cm on alternate trials. Because angular speed was changing by about a factor of 2 from trial to trial, observers had to rely on scaled changes in their surroundings to compensate for the changes in viewing distance. In short, they were being asked to judge objective speed using the transposition scaling first identified by Brown. Despite the guidance provided by feedback, their Weber fractions for transposed speed were considerably less precise than their monocular judgments<sup>8</sup> of angular speed (see the bottom graph in Figure 14.14). Frame effects were not much better than disparity in producing precise judgments of objective speed.

These results argue that speed constancy depends on scaling the distance traversed rather than direct scaling of the angular speed. Otherwise, it is difficult to explain why objective speed thresholds are two to three times the angular thresholds. Based on a propagation-of-error calculation in which different sources of error were assumed to be independent, McKee and Welch (1989) predicted objective speed thresholds from the measured errors in the angular speed and in the depth judgments; the measured objective speed thresholds were significantly higher than their predictions. In contrast, similar predictions for size constancy were in good agreement with the measured thresholds. We thus attribute the loss of precision in objective speed thresholds to the indirect calculation of speed.

In laboratory situations, the traverse has a well-defined, if arbitrary, length. How can speed constancy operate in natural environments where the traverse length for an object in motion is undefined because the object is still moving? Does the observer have to wait until the object disappears before judging its speed? Objects usually have a static background that can be used to scale the distance moved per unit time. Because both the average angle subtended by the background texture and the angular velocity are scaled with viewing distance, the "proportion" of background moved per unit time remains constant as the viewing distance increases – the transposition effect again. Temporal frequency coding, the alternative to the indirect calculation of speed, may underlie speed constancy in some conditions. The temporal frequency spectrum of any transient signal is invariant with distance because the decrease in the angular velocity and the increase in the spatial frequency spectrum cancel. Temporal frequency judgments are somewhat less precise than speed judgments (McKee et al., 1986; Pasternak, 1987; Smith & Edgar, 1991), which would account for the imprecision of speed constancy.

### Three procedures for calculating speed

In summary, angular speed can be calculated with high precision from the signals generated by primary motion units (Grzywacz & Yuille, 1990; Heeger,

### Wide Receiver is moving 12 miles/hour

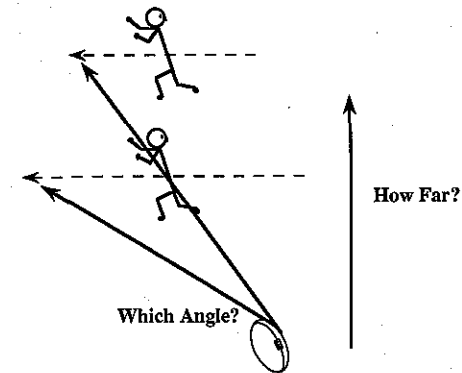


Figure 14.15. Example showing that angular speed information is needed to guide eye and body movements. A quarterback in American football needs information about *angular* speed in order to determine the angle at which the football needs to be thrown to reach the receiver. This angular information is lost in objective speed calculation.

1987). Objective speed, on the contrary, depends on two less precise calculations: (1) temporal frequency coding and (2) the indirect calculation of speed from duration and a scaled representation of distance traversed.

It is nevertheless puzzling that the visual system does not transform the angular speed signal directly into some representation of objective speed. One answer comes from a consideration of the uses of angular speed information. Like angular sizes, angular speeds define the "background" of the visual world; they are the raw data of optic flow, providing information about object boundaries and relative depth. Angular speed also guides the human motor system; it is the primary input to pursuit eye movements and other kinds of movement that involve rapid adjustment in the pursuit of a moving object. In American football, the best quarterbacks<sup>9</sup> adjust their passes so that the ball will arrive a second or two later at the predicted location of a receiver running across the football field. Knowing that the wide receiver is running at a rate of 12 miles an hour is not helpful because objective speed carries no information about the angle required for the ball to reach its target (see Figure 14.15). The quarterback needs separate, independent information about both the receiver's distance and the receiver's *angular* speed so that he can adjust the angle of his arm and hand in a throwing motion that will deliver the ball to the appropriate location. Of course, the quarterback can reconstruct angular speed indirectly from objective speed and the estimated depth of the receiver, but this reconstruction would necessarily have more error than the direct estimate of angular speed because it would include noise from the depth signal. It is a much better strategy to use the angular speed signal, uncontaminated by depth noise, as the basis of fine motor control.



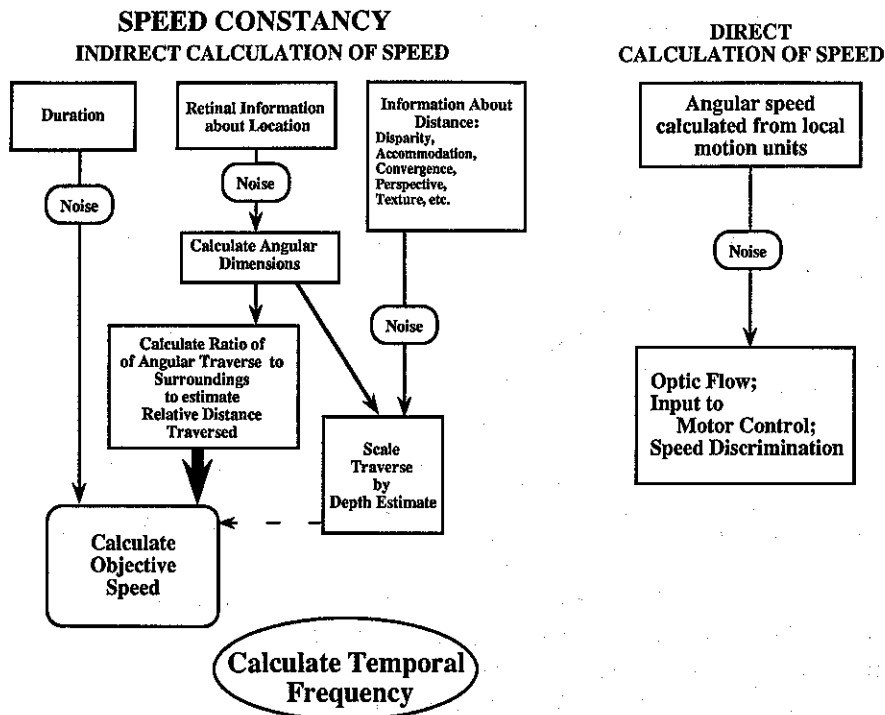


Figure 14.16. Flow charts showing dual calculation of objective and angular speed information. Objective speed can be calculated from temporal frequency information; it can also be calculated indirectly from target duration and from the distance traversed, where traverse is scaled by the same procedures as in size constancy. Angular speed information is based on calculation from signals generated by motion detectors.

Objective speed information is chiefly useful in maintaining an object's identity; the apparent speed of an object moving across our visual field should not change as we change our viewing distance. Whereas body movements require high precision, object identity does not, so there is no functional demand for a highly precise representation of objective speed. In psychophysical experiments, learning fine speed discrimination is quite difficult, which suggests that we are largely unaware of the precise background signal that guides our movements and defines our visual world. When we *attend* to a target, we see its scaled objective speed.

An overview of speed constancy processing is shown in Figure 14.16. As in the case of size constancy, we are proposing a dual calculation of the objective and angular speeds. Wallach contended that speed constancy could not be an extension of size constancy because it did follow the same rules that

governed size constancy. Speed constancy does indeed depend on a different algorithm from the one that governs size constancy, but many of the components are similar. The "distance traversed" is scaled by the same information as size, but relational scaling (large solid arrow) is weighted so heavily that it generally overwhelms depth scaling (small dotted arrow). In addition, the calculation of temporal frequency provides a strong alternative to scaling the distance traversed. There is no evidence indicating how these two alternative procedures for speed constancy are combined, so we have left them separate in the diagram.

### Notes

1. In all experimental measurements of size constancy, observers judge the *relative* size of one object with respect to another. It is difficult to determine if observers also have correct information about the true physical size of the objects (the *absolute* size). Thus, we generally will not distinguish between relative and absolute judgments of objective size.
2. Sedgwick (1986) contains a superb summary on the use of perspective and texture information to estimate distance.
3. Diopter is a measure of lens power and equals the reciprocal of the focal length in meters.
4. The angle subtended at a point by straight lines from the rotation centers of the eyes.
5. Predicted error =  $\sqrt{(\text{error}_1^2 + \text{error}_2^2)}$ .
6. The lower line is about 20% too short.
7. Spatial frequency varies inversely with angular dot size. Viewing the dots at half the distance would double their angular subtense and roughly halve the peak of the spatial frequency spectrum. Changes in the spacing of randomly positioned dots should not affect the content of the spectrum.
8. For the monocular control, observers discriminated small differences in angular speed at each of the two distances (28 cm and 57 cm) while remaining stationary; their Weber fractions for these two fixed distances were averaged.
9. The quarterback is the player who throws the football forward to the receivers after it is handed to him by the player in the center of the forward line.

### References

- Andrews, D. P. (1964). Error-correcting perceptual mechanisms. *Quarterly Journal of Experimental Psychology*, 16, 105-115.
- Bevington, P. R. (1969). *Data reduction and error analysis for the physical sciences*. New York: McGraw-Hill.
- Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*, (pp. 288-299). New York: D. Appleton-Century.
- Boring, E. G. (1946). The perception of objects. *American Journal of Physics*, 14, 99-107.
- Brown, J. F. (1931). The visual perception of velocity. *Psychologische Forschung*, 14, 199-232.
- Burbeck, C. A. (1987a). Position and spatial frequency in large-scale localization judgments. *Vision Research*, 27, 417-427.
- Burbeck, C. A. (1987b). On the locus of spatial frequency discrimination. *Journal of the Optical Society of America A*, 4, 1807-1813.
- Campbell, F. W. (1957). The depth of field of the human eye. *Optica Acta*, 4, 157-164.

- Carlson, V. R. (1960). Overestimation in size-constancy judgments. *American Journal of Psychology*, 73, 199-213.
- Chen, Y., Bedell, H. E., and Frishman, L. J. (1995). Velocity discrimination between stimuli of different spatial frequencies. *Investigative Ophthalmology and Visual Science (Supplement)*, 36, S54.
- Coren, S., & Girgus, J. S. (1978). *Seeing is deceiving: The psychology of visual illusions*. Hillsdale, NJ: Erlbaum.
- Diener, H. C., Wist, E. R., Dichgans, J., & Brandt, Th. (1976). The spatial frequency effect on perceived velocity. *Vision Research*, 16, 169-176.
- Epstein, W. (1967). The assumptive context: I. The perception of size and distance. In *Varieties of perceptual learning* (pp. 21-53). New York: McGraw-Hill.
- Epstein, W. (1973). The process of 'taking-into-account' in visual perception. *Perception*, 2, 267-285.
- Epstein, W. (1978). Two factors in the perception of velocity at a distance. *Perception & Psychophysics*, 24, 105-114.
- Epstein, W., & Broota, K. D. (1986). Automatic and attentional components in perception of size-at-a-distance. *Perception & Psychophysics*, 40, 256-262.
- Epstein, W., & Cody, W. J. (1980). Perception of relative velocity: A revision of the hypothesis of relational determination. *Perception*, 9, 47-60.
- Epstein, W., Park, J., & Casey, A. (1961). The current status of the size-distance hypothesis. *Psychological Bulletin*, 58, 491-514.
- Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, 87, 411-434.
- Franklin, S. S., Ross, H. E., & Weltman, G. (1970). Size-distance adaptation invariance in perceptual adaptation. *Psychonomic Science*, 21, 229-231.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Gilinsky, A. S. (1955). The effect of attitude upon the perception of size. *American Journal of Psychology*, 68, 173-192.
- Gillam, B. (1973). The nature of size scaling in the Ponzo and related illusions. *Perception & Psychophysics*, 14, 353-357.
- Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature (London)*, 199, 678-680.
- Gregory, R. L. (1973). The confounded eye. In R. L. Gregory & E. H. Gombrich (Eds.), *Illusion in nature and art* (pp. 12-41). New York: Scribner's.
- Gregory, R. L. (1987). Emmert's law. In R. L. Gregory (Ed.), *The Oxford companion to the mind* (pp. 218-219). Oxford: Oxford University Press.
- Gregory, R. L., Wallace, J. G., & Campbell, F. W. (1959). Changes in the size and shape of visual after-images observed in complete darkness during changes of position in space. *Quarterly Journal of Experimental Psychology*, 11, 54-56.
- Grzywacz, N. M., & Yuille, A. L. (1990). A model of the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society London B*, 239, 129-161.
- Harvey, L. O., & Leibowitz, H. W. (1967). Effects of exposure duration, cue reduction, and temporary monocularly on size matching at short distances. *Journal of the Optical Society of America*, 57, 249-253.
- Heeger, D. J. (1987). A model for the extraction of image flow. *Journal of the Optical Society of America A*, 4, 1455-1471.
- Hell, W. (1978). Movement parallax: An asymptotic function of amplitude and velocity of head motion. *Vision Research*, 18, 629-635.
- Hochberg, J. (1972). Perception II. Space and movement. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth & Schlosberg's experimental psychology* (3rd ed., pp. 475-550). New York: Holt, Rinehart & Winston.
- Holway, A. H., & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology*, 54, 21-37.

- Huber, J., & Davies, I. (1995). Motion parallax: A weak cue for depth in telepresence systems. *Perception (Supplement)*, 24, 106.
- Humphrey, N. K., & Morgan, M. J. (1965). Constancy and the geometric illusions. *Nature*, 206, 744-745.
- Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, 31, 1351-1360.
- Klein, S. A., & Levi, D. M. (1987). Position sense of the peripheral retina. *Journal of the Optical Society of America A*, 4, 1543-1553.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389-412.
- Legge, G. E., Mullen, K. T., Woo, G. C., & Campbell, F. (1987). Tolerance to visual defocus. *Journal of the Optical Society of America A*, 4, 851-863.
- Leibowitz, H. W., & Harvey, L. O., Jr. (1967). Size matching as a function of instructions in a naturalistic environment. *Journal of Experimental Psychology*, 74, 378-382.
- Lichten, W., & Lurie, S. (1950). A new technique for the study of perceived size. *American Journal of Psychology*, 63, 280-282.
- Mandriota, F. J., Mintz, D. E., & Notterman, J. M. (1962). Visual velocity discrimination: Effects of spatial and temporal cues. *Science*, 138, 437-438.
- McKee, S. P. (1981). A local mechanism for differential velocity detection. *Vision Research*, 21, 491-500.
- McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research*, 30, 1763-1779.
- McKee, S. P., Silverman, G. S., & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research*, 26, 609-620.
- McKee, S. P., & Welch, L. (1989). Is there a constancy for velocity? *Vision Research*, 29, 553-561.
- McKee, S. P., & Welch, L. (1992). The precision of size constancy. *Vision Research*, 32, 1447-1460.
- McKee, S. P., Welch, L., Taylor, D. G., & Bowne, S. F. (1990). Finding the common bond: Stereocoactivity and the other hyperacuties. *Vision Research*, 30, 879-891.
- Morgan, M. J. (1989). Vision of solid objects. *Nature*, 339, 101-103.
- Morgan, M. J. (1992). On the scaling of size judgments by orientational cues. *Vision Research*, 32, 1433-1445.
- Nakayama, K. (1994). James J. Gibson - an appreciation. *Psychological Review*, 101, 329-335.
- Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of 3-D length. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 173-186.
- Orban, G. A., de Wolf, J., & Maes, H. (1984). Factors influencing velocity coding in the human visual system. *Vision Research*, 24, 33-39.
- Over, R. (1960). The effect of instructions on size-judgments under reduction conditions. *American Journal of Psychology*, 73, 599-602.
- Pasternak, T. (1987). Discrimination of differences in speed and flicker rate depends on directionally-selective mechanisms. *Vision Research*, 27, 1881-1890.
- Rock, I. (1977). In defense of unconscious inference. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes* (pp. 95-125). New York: Wiley.
- Rock, I., & Ebenholtz, S. (1959). The relational determination of perceived size. *Psychological Review*, 66, 387-401.
- Rock, I., Hill, A. L., & Fineman, M. (1968). Speed constancy as a function of size constancy. *Perception & Psychophysics*, 4, 37-40.
- Rogers, B. J., & Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology*, 41A, 697-717.
- Rogers, B. J., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, 8, 125-134.

- Ross, H. E., Franklin, S. S., Weltman, G., & Lennie, P. (1970). Adaptation of divers to size distortion under water. *British Journal of Psychology*, *61*, 365-373.
- Sedgwick, H. A. (186). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. I, pp. 21-57). New York: Wiley.
- Sedgwick, H. A., & Nicholls, A. L. (1993). Interaction between surface and depth in the Ponzo illusion. *Investigative Ophthalmology and Visual Science Supplement*, *34*, 1184.
- Smith, A. T., & Edgar, G. K. (1991). The separability of temporal frequency and velocity. *Vision Research*, *31*, 321-326.
- Smith, O. W., & Sherlock, L. (1957). A new explanation of the velocity-transposition phenomenon. *American Journal of Psychology*, *70*, 102-105.
- Templeton, W. B., & Green, F. A. (1968). Chance results in utrocular discrimination. *Quarterly Journal of Experimental Psychology*, *20*, 200-203.
- Thouless, R. H. (1931). Phenomenal regression to the real object. *British Journal of Psychology*, *21*, 335-359.
- Wallach, H. (1939). On constancy of visual speed. *Psychological Review*, *46*, 541-552.
- Wenderoth, P. (1976). The contribution of relational factors to line-length matches. *Perception*, *5*, 265-278.
- Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*, 2685-2696.
- Yuille, A. L., & Bülthoff, H. H. (1993). *Bayesian decision theory and psychophysics*. CogSci Memo No. 20. Tübingen: Max-Planck-Institut für biologische Kybernetik, Arbeitsgruppe Bülthoff.
- Zohary, E., & Sittig, A. C. (1993). Mechanisms of velocity constancy. *Vision Research*, *33*, 2467-2478.