

Range image statistics can explain the anomalous perception of length

Catherine Q. Howe and Dale Purves*

Department of Neurobiology, Box 3209, Duke University Medical Center, Durham, NC 27710

Contributed by Dale Purves, August 7, 2002

A long-standing puzzle in visual perception is that the apparent extent of a spatial interval (e.g., the distance between two points or the length of a line) does not simply accord with the length of the stimulus but varies as a function of orientation in the retinal image. Here, we show that this anomaly can be explained by the statistical relationship between the length of retinal projections and the length of their real-world sources. Using a laser range scanner, we acquired a database of natural images that included the three-dimensional location of every point in the scenes. An analysis of these range images showed that the average length of a physical interval in three-dimensional space changes systematically as a function of the orientation of the corresponding interval in the projected image, the variation being in good agreement with perceived length. This evidence implies that the perception of visual space is determined by the probability distribution of the possible real-world sources of retinal images.

As the orientation of a linear stimulus in the retinal image changes, so does its apparent length. Thus, a line that projects vertically appears to be longer than the same line presented horizontally, the maximum length being seen when the stimulus is oriented 20–30° from the vertical axis (refs. 1–4; Fig. 1). This variation is evidently a particular manifestation of the general tendency to perceive the extent of any spatial interval differently as a function of its orientation in the retinal image. For instance, the apparent distance between a pair of dots varies systematically with the orientation of the imaginary line between them (5), and a perfect square or circle appears to be slightly elongated along its vertical axis (6, 7). Despite extensive study of these phenomena during the past 150 years (8–19), neither a quantitative explanation of these effects nor a generally accepted biological rationale has been forthcoming.

The explanation we have examined here is that the variable perception of length as a function of stimulus orientation represents a solution to the problem presented by the inevitable ambiguity of retinal projections, namely that retinal images cannot uniquely specify their physical sources (20). Thus, a given line in the retinal image could have been generated by an infinite number of real-world objects with different lengths, located at different distances and in different 3D orientations. The physical source of a retinal stimulus nonetheless is what the observer must respond to with appropriate visually guided behavior. One solution to this dilemma would be to generate percepts according to the probability distribution of the possible physical sources of the retinal stimulus. To test this hypothesis with respect to the perception of visual space, the arrangement of objects in the physical world must be related to their projected images. Accordingly, we used a laser range scanner (21, 22) to acquire a database of natural scenes that included the location in 3D space of every pixel in the images (Fig. 2). We then could relate any given projection to its real-world sources, in this way asking whether the probabilistic relationship between the length of intervals in images and the length of their physical sources can explain the anomalous perception of length as a function of stimulus orientation.

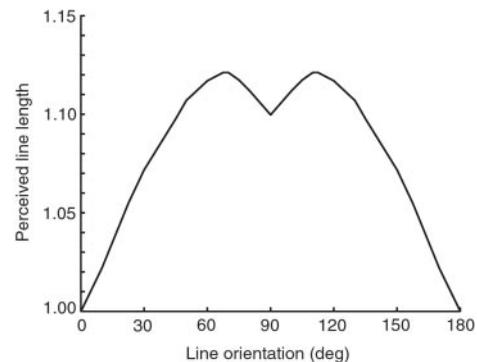


Fig. 1. Variation in the apparent length of a linear stimulus as a function of its orientation in the retinal image. The function shown is an average of the psychophysical data reported in refs. 2–4. Orientation is expressed as the angle between the line and the horizontal axis. The maximum length seen by observers occurs when the line is oriented 20–30° from the vertical axis, at which point it appears about 10–15% longer than the minimum length seen when the orientation of the stimulus is horizontal.

Materials and Methods

Acquiring the Range Image Database. Range images were acquired by using an LMS-Z210 3D laser scanner (Riegl, Orlando, FL) controlled by Riegl 3-D-RISCAN software installed on a laptop computer. The scanner combines a laser range-finder with a true color channel, thus providing digitized images with accurate distance as well as luminance information for every pixel in the scene. The range-finding performance of the system is from 2 to approximately 300 m, with an accuracy of ± 25 mm over this full range. Twenty-five wide-field images of fully natural scenes covering a 333° [horizontal (H)] \times 80° [vertical (V)] field of view with an angular resolution of 0.144° were acquired in the Sarah P. Duke Gardens on the Duke University campus and the nearby Duke Forest. The scanner system was mounted on a surveyor's tripod such that the origin of the laser beam was at a height of 165 cm; the apparatus was leveled in the horizontal plane before acquiring each image.

Each of these wide-field range images comprising the 3D locations of all of the imaged points in a spherical polar coordinate system was transformed into a series of ≈ 600 2D projections corresponding to a 20° \times 20° field of view. This transformation was carried out by placing an imaginary projection plane at the origin of the polar coordinate system (i.e., the origin of the laser beam) and altering the orientation of the plane progressively in steps of 5° in both azimuth and elevation. A region of the 3D world represented by the range image directly in front of the imaginary plane then was projected onto the plane by using a pinhole model, which provides a good approximation of the retinal image formation process (23, 24). The result was a series of 2D images measuring approximately 140 \times 140 pixels.

Abbreviations: l , projected interval length; θ , projected interval orientation; λ , ratio of physical length to projected length; ϕ , inclination in depth.

*To whom reprint requests should be addressed. E-mail: purves@neuro.duke.edu.

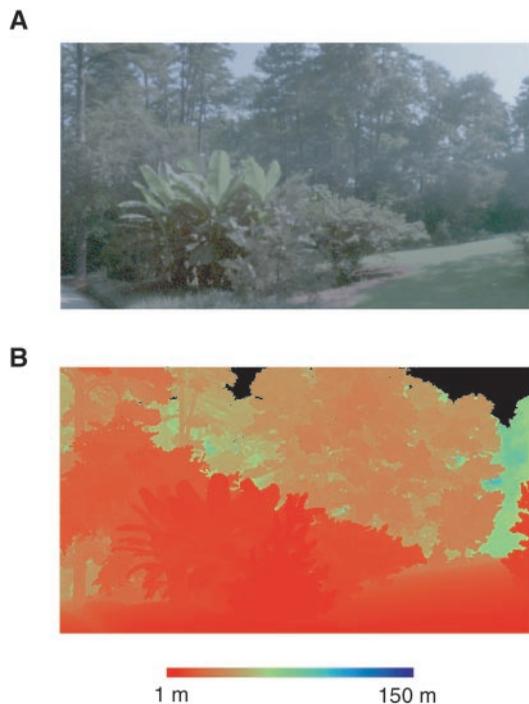


Fig. 2. A representative image acquired by the laser range scanner (only a small portion of the full, wide-field image is shown). (A) True color image generated by the scanner. (B) Corresponding range image; the physical distance of each point in the scene from the origin of laser beam in the scanner is indicated by color-coding. Black areas (the sky) are the points in the scene from which no laser reflection was recorded; such points were omitted from the analysis.

At the same time, the 3D locations of the pixels in each projected image were transformed into coordinates in a Cartesian system whose XY plane was parallel to the image projection plane. The end result of this procedure was a database of $\approx 15,000$ different 2D image projections, together with the 3D coordinates in Cartesian space of each constituent pixel, thus representing a full range of geometrical relationships between the projected image and the real world.

Sampling and Analyzing Spatial Intervals Between Pairs of Points.

Pairs of points were selected randomly in these 15,000 images. The first point of a pair was taken from the central region of the image within a circular area whose diameter was half of the image width; the second point then was sampled within a circular area of the same size centered at the first point. This method was adopted to avoid the effect of the image boundary on sampling (25). Because the size of the images from which the spatial intervals were sampled was approximately 140×140 pixels (see above), the length of the sampled intervals ranged from 1 to 35 pixels (i.e., up to one-fourth of 140).

We then analyzed the frequency distribution of the ratio λ (physical length-to-projected length ratio) as a function of the projected interval orientation (θ ; bin width = 1°); the distribution as a function of the projected interval length (l ; bin width = 1 pixel) also was examined. The reason for this latter assessment was that, as a consequence of the discrete composition of digital images, intervals sampled at certain values of θ inevitably have large l values. Because λ generally decreases as l increases (see Fig. 3B), the average λ at these values of θ would have been artificially small if the frequency distributions of λ were simply integrated over different values of l . Thus, to obtain a better indication of the variation of λ as a function of θ , we removed the

influence of l by normalization. The normalization was carried out by calculating the ratio of the mean of λ at each θ to the mean of λ at 0° for each value of l and then averaging the ratios across different values of l .

Sampling and Analyzing Contours. Luminance edges (i.e., pixels that fall along luminance boundaries) were extracted from the images by using the method described by Canny (26); the magnitude and orientation of luminance gradients across the edges were determined by using the steerable filters developed and described by Freeman and Adelson (27) and improved by Yu *et al.* (28). Collinear edge elements then were grouped into straight line segments by using the algorithm described by Sarkar and Boyer (29). Only lines greater than 8 pixels in length were included in the analysis, primarily because lines of this length or greater are readily seen as straight lines in the images. Most luminance edges that qualify as straight lines in the analysis lie either in the ground plane or in the more rectilinear components of the scenes (e.g., fewer straight lines derive from leaves than tree trunks; see Fig. 4A).

Results

The frequency distribution of intervals in physical space corresponding to a given interval in the projected images was determined by sampling intervals between random pairs of points in the image database and by sampling line segments associated with luminance contrast boundaries. The rationale for the first method is that the perception of length is pertinent to all types of spatial intervals (e.g., the distances between two points, the length of a line, or the dimensions of more complex geometries). Sampling intervals between random points thus has the advantage of taking into account all these categories of spatial information. We also examined luminance contrast boundaries because visual images often are studied in these terms, despite the fact that edges represent only a small fraction of the spatial intervals routinely experienced (thus, the contour data are essentially a subset of the data derived from the point-pair analysis). In both approaches, the physical length of the interval in 3D space was calculated from the range information in the database. We then divided the length of the physical interval (L) by the length of the corresponding interval in the projected image (l) to obtain λ , thus relating the projections to their physical sources. By sampling a large number of spatial intervals (>250 million in the point pair analysis, and $\approx 500,000$ in the contour analysis), a frequency distribution of this ratio (λ) was generated for all possible projected interval orientations (θ) in the images (Figs. 3 and 4).

It is apparent in Figs. 3B and C and 4B that the frequency distribution of λ varies systematically as a function of θ (the function in Fig. 4B is much noisier because of the relatively small sample size of contours; the presentation that follows therefore is based primarily on the point-pair data). The maxima of the functions in Figs. 3C and 4B occur when θ is $20\text{--}30^\circ$ from the vertical. Thus, the average length of real-world spatial intervals underlying retinal projections changes continually as a function of the orientation of the projections, being greatest when the projection is near, but not at, vertical. The magnitude of the variation (from minimum to maximum) in Fig. 3C, which takes into account all categories of spatial intervals, is about 15%, in good agreement with the psychophysical data shown in Fig. 1. These results support the hypothesis that the anomalous perception of length as a function of projected orientation is explained by the systematically different average length of the generative physical sources.

We next asked which aspect of the arrangement of real-world objects is responsible for the variation of the ratio (λ) of the physical length to projected length as a function of the projected orientation. Two factors can affect this ratio: (i) the distance of

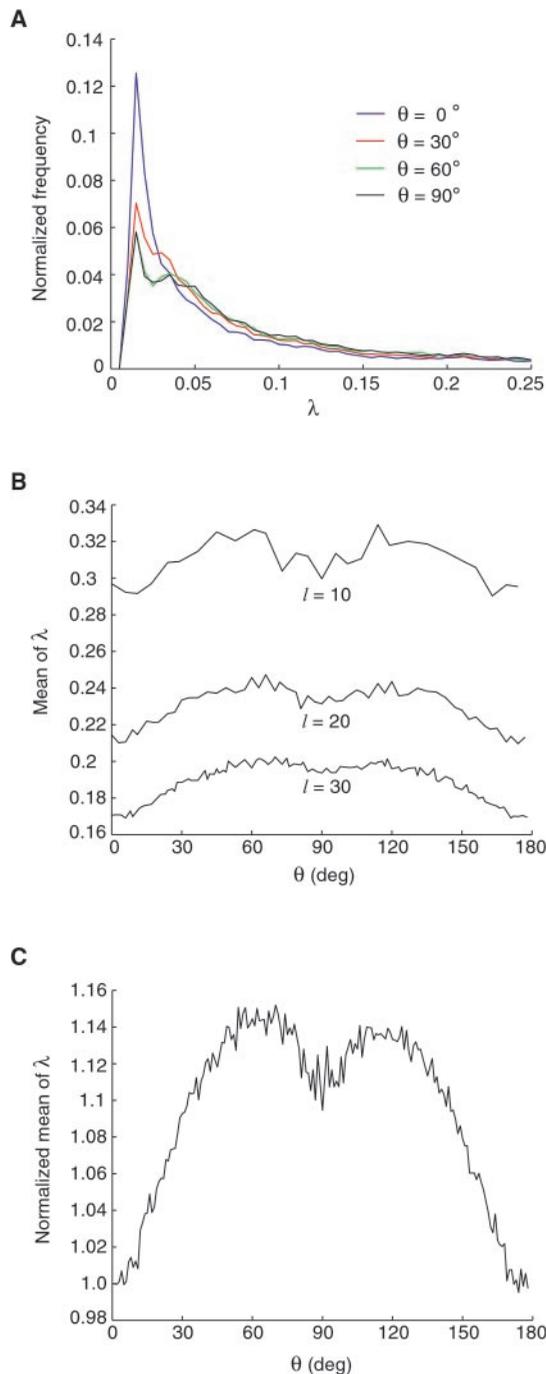


Fig. 3. Statistical relationship between the spatial intervals in the projected images and the corresponding physical intervals in 3D space. (A) Examples of the frequency distribution of the physical length to projected length ratio (λ , in meters/pixel) at four different θ values. The l in all three examples is 20 pixels. (B) The mean value of λ plotted as a function of the θ . Examples of the function obtained at three different l values (in pixels) are shown (see *Materials and Methods*). (C) Normalized mean of λ averaged across all values of projected interval length plotted as a function of θ . The similarity of this curve to the psychophysical function in Fig. 1 indicates that the statistical relationship between projected length and physical length accords with the perceptual variation of apparent line length.

physical intervals from the image plane and (ii) the inclination of the physical intervals in depth (Fig. 5A). With respect to the first of these possible influences, we found little difference between the average distance from the image plane of vertically

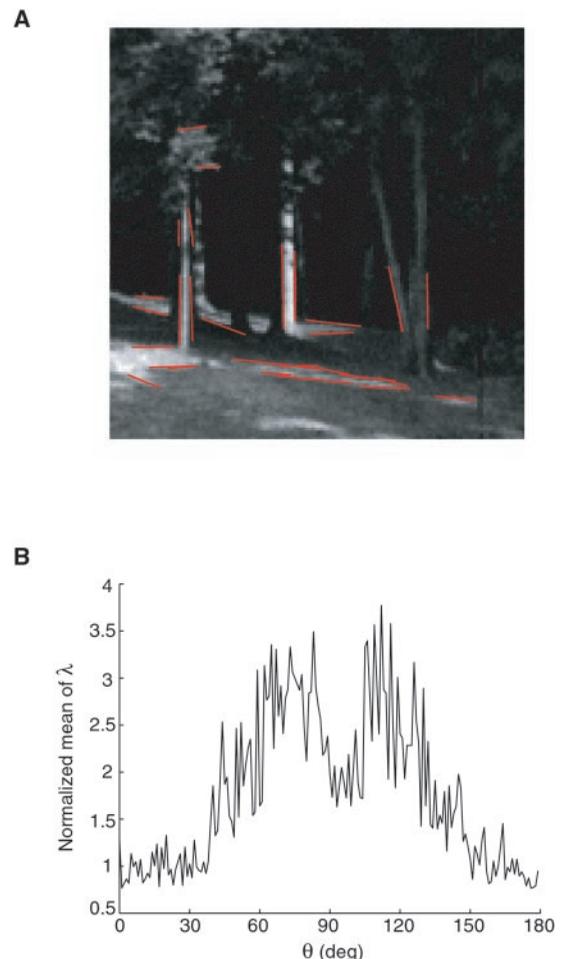


Fig. 4. Statistical relationship between the luminance boundaries in the projected images and the corresponding physical sources in 3D space. (A) Example of straight-line segments (shown in red) extracted from a scene in the database. The images were converted to gray scale before the extraction of contours. (B) The normalized mean of the physical-to-projected length ratio (λ ; in meters/pixel) plotted as a function of θ of luminance contours in the image plane (bin width of $\theta = 1^\circ$). The data obtained from this analysis, which represent a small subset of the data shown in Fig. 3, show the same peaks approximately $20\text{--}30^\circ$ from the vertical and the same trough at 90° . The variation from the minimum to the maximum of this function is greater compared with the function in Fig. 3C. This greater variation arises because the extracted line segments tend to be in the ground plane (see Fig. 6 B and C for the effect of the ground plane) or in the more rectilinear components of the scenes (see A).

and horizontally projecting intervals (about 1.5%, the average distance of the physical sources of horizontal intervals being slightly greater than sources of vertical ones). Regarding the second possibility, inclination in depth (ϕ) is the angle of a line, or an imaginary line, with respect to the frontal (i.e., image) plane (if $\phi = 0^\circ$, the interval is in a frontal plane). Thus, the larger the angle ϕ , the larger the ratio of the physical interval to its projected interval. The mean of the frequency distribution of ϕ as a function of the projected interval orientation (θ) is shown in Fig. 5B. The similar variation of ϕ and λ as functions of θ (compare Figs. 3C and 5B) indicates that the primary reason why the average ratio of physical length to projected length is greater for vertical or near-vertical intervals in natural images is that the physical sources of such intervals tend to be more inclined away from the frontal plane than the sources of intervals in other orientations.

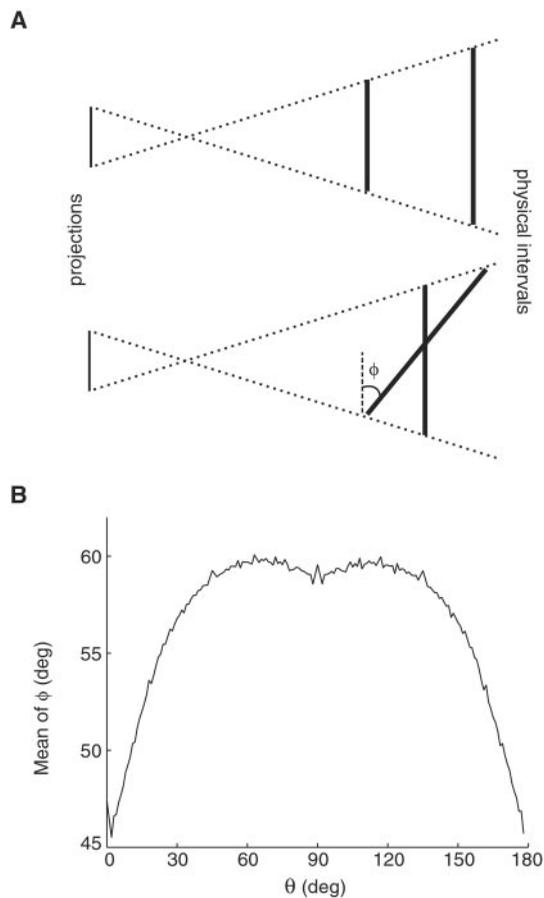


Fig. 5. The influence of distance and 3D orientation on the ratio (λ) of physical length to projected length. (A) Diagram of these effects. The greater the distance of the physical source from the projection plane and the more the source inclines away from the frontal plane (i.e., the larger ϕ), the larger the ratio λ . (B) The mean ϕ of physical intervals as a function of θ .

Why, then, do the physical sources of vertical or near-vertical projections tend to be more inclined in depth? One possibility is that this bias is caused by the presence of the ground plane in most natural scenes. The more a physical interval extends in depth in the ground plane, the more its projection will tend toward the vertical axis (Fig. 6A). This geometrical fact, coupled with the prevalence of the ground plane, would cause the physical sources of vertical intervals, on average, to incline more in depth and, thus, to be longer compared with the sources of horizontal intervals. To test this possibility, we examined intervals from two subsets of images in the database, one in which the ground plane was lacking and the other in which the ground plane was the predominant component. In the sample lacking the ground plane, the difference in the inclination in depth of the physical sources of vertical and horizontal intervals was diminished compared with the database as a whole; in contrast, this difference was exaggerated in the sample containing the ground plane (Fig. 6B). As a result, the ratio of physical length to projected length showed the same pattern of variation (Fig. 6C). These findings indicate that the variation in the statistical relationship between intervals in the images and their physical sources as a function of orientation indeed is caused by this underlying asymmetry of natural scene geometry.

Finally, the peaks near 45° and 135° in the sample lacking the ground plane (see Fig. 6B and C) need to be explained. These peaks presumably are caused by the prevalence of objects in the natural world that are perpendicular or parallel to the ground

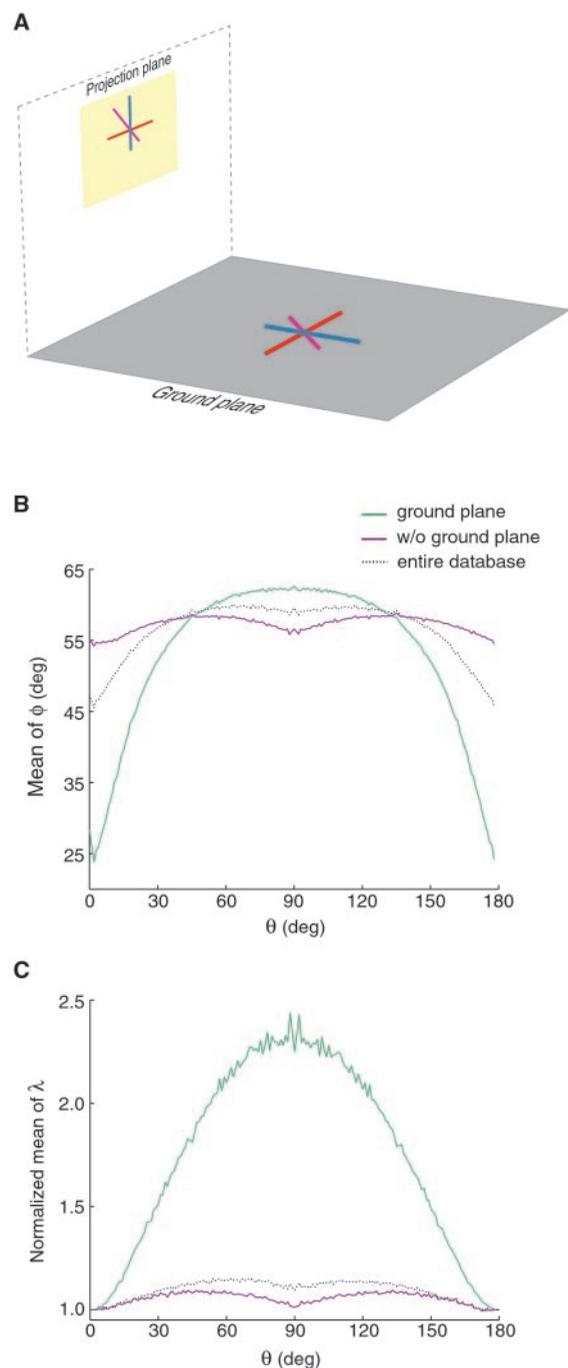


Fig. 6. The contribution of the ground plane to the distribution of the physical sources of differently oriented intervals. (A) Physical intervals on the ground plane (colored lines) that extend further and further away from the image plane will project more and more vertically (see corresponding colored lines). (B) Intervals were sampled from two distinct subsets of the image database, one without the ground plane and the other with the ground plane being the predominant component. The mean ϕ of the physical intervals in each of these samples is plotted as a function of θ . The function obtained from the entire image database (Fig. 5B) is included for comparison. (C) The mean of the physical length-to-projected length ratio (λ) of the intervals in each of the samples plotted as a function of θ .

plane [i.e., the demonstrated predominance of objects in the cardinal axes (30, 31)]. As a result of this bias, vertical and horizontal projections would be less likely to be generated by objects extending in depth compared with oblique projections,

thus accounting for the peaks of average inclination in depth at oblique angles. To examine this possibility, we generated two hypothetical 3D spaces, one that was spherical, and the other rectilinear. Both spaces then were populated with randomly distributed points, and the spatial intervals between pairs of these points projected onto an imaginary plane. In the spherical “world,” the average inclination in depth of the physical intervals was the same for all projected interval orientations. In the rectilinear “world,” however, the average inclination in depth showed distinct peaks at 45° and 135° (data not shown).

Discussion

The discrepancy between the measured length of a spatial interval and its perception has been rationalized in several different ways in the past, including asymmetries in the anatomy of eye (8, 10, 11, 17), the ergonomics of eye movements (5, 32), and cognitive compensation for the foreshortening of vertical lines (12, 14–16). In the last of these theories, which is the one most often cited, vertical lines in the image plane are assumed to be objects on the ground plane that extend in depth; horizontal lines, on the other hand, are taken to be objects parallel to the frontal plane. This general explanation, however, fails to recognize that both vertical and horizontal lines, or lines in any orientation in the image plane, can be generated by physical sources that have any degree of inclination in depth. Thus, theories of this sort do not explain the psychophysical results shown in Fig. 1. Perhaps the most sophisticated approach to date is Craven’s analysis of “zero-crossings” in 2D natural images (4). Because the density of contrast transitions (zero crossings) in filtered images was found to be greater along the vertical lines than lines at other orientations, it was proposed that the visual system calibrates perceived length according to this metric. Although we do not doubt the accuracy of this analysis, there is no obvious reason why the visual system should carry out a computation of this sort.

In contrast, the evidence presented here points to an explanation that is both simple and biologically principled. The physical sources underlying linear projections (or, indeed, any image projection) are deeply uncertain. Thus, the strategy of vision that best can ensure appropriate visually guided behaviors

in response to retinal stimuli of uncertain provenance would be to generate percepts according to the probability distributions of the possible sources. Given the statistical analysis reported here, a vertical interval in the retinal image is seen as longer than the same interval oriented horizontally because its possible real-world sources are, on average, physically longer.

The perceived length of intervals as a function of their projected orientation (see Fig. 1) agrees remarkably well with the probability distribution of the possible stimulus sources when projected orientation is the only consideration (see Fig. 3C). Although this good correlation might be considered a result of the fact that the anomalous perception of length typically has been studied in laboratory settings in which the stimuli impose few constraints on the relevant probability distributions, similar differences in perceived length are apparent in natural settings (18). Evidently, the same ambiguity regarding possible sources exists in a wide variety of circumstances, despite the fact that the statistical relationship between image and source is constrained by different variables.

A number of recent studies of visual perception have examined the statistics of natural images (reviewed in ref. 33). Much of this work has been motivated by the notion that the goal of visual perception is to encode image features with optimal efficiency (34–36) and, therefore, has focused on the statistics of features within the image plane. The approach we have taken here is fundamentally different in that we have explored the statistical relationship between elements in the projected image and the sources of those elements in the real world. Understanding the statistical relationship between natural images and their sources has the potential to explain a wide range of perceptual phenomena and could provide a novel framework for considering the functional significance of the relevant visual cortical circuitry.

We thank Zhiyong Yang and Fuhui Long for assistance in acquiring the image database and for providing helpful suggestions during the course of this work. We also thank David Fitzpatrick, Surajit Nundy, David Schwartz, Sidney Simon, and James Voyvodic for useful comments on the manuscript. This work was supported by National Institutes of Health Grant 29187. C.Q.H. is a Howard Hughes Medical Institute Predoctoral Fellow.

- Shipley, W. C., Mann, B. M. & Penfield, M. J. (1949) *J. Exp. Psychol.* **39**, 548–551.
- Pollock, W. T. & Chapanis, A. (1952) *Q. J. Exp. Psychol.* **4**, 170–178.
- Cormack, E. O. & Cormack, R. H. (1974) *Percept. Psychophys.* **16**, 208–212.
- Craven, B. J. (1993) *Proc. R. Soc. London Ser. B Biol. Sci.* **253**, 101–106.
- Wundt, W. (1862) *Beiträge zur Theorie der Sinneswahrnehmung* (C. F. Winter’sche Verlagshandlung, Leipzig and Heidelberg).
- Sleight, R. B. & Austin, T. R. (1952) *J. Psychol.* **33**, 279–287.
- McManus, I. C. (1978) *Br. J. Psychol.* **69**, 369–370.
- Kuennapas, T. M. (1957) *J. Exp. Psychol.* **53**, 405–407.
- Avery, G. C. & Day, R. H. (1969) *J. Exp. Psychol.* **81**, 376–380.
- Pearce, D. & Matin, L. (1969) *Percept. Psychophys.* **6**, 241–243.
- Restle, F. & Merryman, C. (1969) *J. Exp. Psychol.* **81**, 297–302.
- Gregory, R. L. (1974) *Concepts and Mechanisms of Perception* (Duckworth, London).
- Thompson, J. G. & Schiffman, H. R. (1974) *Vision Res.* **14**, 1463–1465.
- Girgus, J. S. & Coren, S. (1975) *Can. J. Psychol.* **29**, 59–65.
- Schiffman, H. R. & Thompson, J. G. (1975) *Perception* **4**, 79–83.
- von Collani, G. (1985) *Percept. Mot. Skills* **61**, 523–531.
- Prinzmetal, W. & Gettleman, L. (1993) *Percept. Psychophys.* **53**, 81–88.
- Higashiyama, A. (1996) *Percept. Psychophys.* **58**, 259–270.
- Robinson, J. O. (1998) *The Psychology of Visual Illusion* (Dover, New York).
- Knill, D. C. & Richards, W. (1996) *Perception as Bayesian Inference* (Cambridge Univ. Press, Cambridge, U.K.).
- Besl, P. J. (1988) *Mach. Vision Appl.* **1**, 127–152.
- Maatta, K., Kostamovaara, J. & Myllyla, R. (1993) *Appl. Optics* **32**, 5334–5347.
- Palmer, S. E. (1999) *Vision Science: Photons to Phenomenology* (MIT Press, Cambridge, MA), p. 24.
- Rodieck, R. W. (1998) *The First Steps in Seeing* (Sinauer, Sunderland, MA), p. 22.
- Binder, K. (1986) *Monte Carlo Methods in Statistical Physics* (Springer, Berlin).
- Canny, J. (1986) *IEEE Trans. Pattern Anal. Machine Intell.* **8**, 679–698.
- Freeman, W. T. & Adelson, E. H. (1991) *IEEE Trans. Pattern Anal. Machine Intell.* **13**, 891–906.
- Yu, W., Daniilidis, K. & Sommer, G. (2001) *IEEE Trans. Image Processing* **10**, 193–205.
- Sarkar, S. & Boyer, K. L. (1994) *IEEE Trans. Syst. Man. Cybern.* **24**, 246–267.
- Coppola, D. M., Purves, H. R., McCoy, A. N. & Purves, D. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 4002–4006.
- Switkes, E., Mayer, M. J. & Sloan, J. A. (1978) *Vision Res.* **18**, 1393–1399.
- Luckiesh, M. (1922) *Visual Illusions: Their Causes, Characteristics and Applications* (Van Nostrand Reinhold, New York).
- Simoncelli, E. P. & Olshausen, B. A. (2001) *Annu. Rev. Neurosci.* **24**, 1193–1216.
- Attneave, F. (1954) *Psychol. Rev.* **61**, 183–193.
- Barlow, H. B. (1961) in *Sensory Communication*, ed. Rosenblith, W. A. (MIT Press, Cambridge, MA), pp. 217–234.
- Field, D. J. (1994) *Neural Comput.* **6**, 559–601.