

Why pictures look right when viewed from the wrong place

Dhanraj Vishwanath¹, Ahna R Girshick¹ & Martin S Banks^{1,2}

A picture viewed from its center of projection generates the same retinal image as the original scene, so the viewer perceives the scene correctly. When a picture is viewed from other locations, the retinal image specifies a different scene, but we normally do not notice the changes. We investigated the mechanism underlying this perceptual invariance by studying the perceived shapes of pictured objects viewed from various locations. We also manipulated information about the orientation of the picture surface. When binocular information for surface orientation was available, perceived shape was nearly invariant across a wide range of viewing angles. By varying the projection angle and the position of a stimulus in the picture, we found that invariance is achieved through an estimate of local surface orientation, not from geometric information in the picture. We present a model that explains invariance and other phenomena (such as perceived distortions in wide-angle pictures).

Pictures have widespread use because, in the convenient format of a two-dimensional (2D) surface, they allow viewers to perceive three-dimensional (3D) scene information. Their usefulness stems, in large part, from the fact that the viewer's eye need not be at the geometrically correct location—the center of projection (CoP)—to create an acceptable impression of the scene. Painters^{1–5}, photographers^{6–9}, cinematographers¹⁰, computer scientists^{11–13} and vision scientists^{14–19} have for years wondered how the perceptual invariance across viewing position is achieved, and this is the central question considered here.

Figure 1a–c demonstrates that the perceived shape of pictured objects is largely invariant with changes in viewing position. It shows that perceived shape is determined by more than the pattern of light striking the eye; it is also affected by the sensed orientation of the picture surface. Here we examine the visual mechanisms underlying these phenomena. Do the mechanisms rely on the geometry of the depicted scene? Or do they rely simply on the orientation of the picture surface?

Before addressing these questions, we explain the principle for creating pictures: perspective projection. **Figure 2a–e** illustrates the projection of a 3D scene onto a picture or projection plane, and its subsequent projection onto the retina. Light rays from the scene are projected toward the CoP, creating the light field²⁰ or optic array²¹. The picture is the intersection of the light field with the projection plane. The dimensions of the image of an object in the picture are affected by two factors: projective scaling (expansion as the distance from the CoP to the projection plane increases) and projective stretching (lengthening in the direction of the slant of the surface). For a sphere stretching is indicated by the difference between the horizontal and vertical

dimensions of its image in the projection plane:

$$\begin{aligned} a &\approx A(p/d)/\cos(S) \\ b &\approx B(p/d) \end{aligned} \quad (1)$$

where d and p are the distances from the CoP to the sphere and the projection plane, respectively, and S is the slant of the plane at the point of interest. Equations (1) become exact as the size of the region in the picture approaches zero²². The scaling effect (p/d) applies equally to a and b , whereas the stretching ($1/\cos(S)$) applies only to a . To a first approximation, the shape of the image of a sphere is affected only by stretching; for an object such as a slanted plane, however, the left and right edges of the image will be scaled differently in the projection plane, and thus both scaling and stretching come into play.

When an observer views the picture, there is a second projection from the picture to the retina. This causes two effects: scaling and foreshortening. For the picture of a sphere, the angular dimensions of the image at the retina are (**Fig. 2e**):

$$\begin{aligned} \alpha &\approx a(f/v)\cos(S_{\text{local}}) \\ \beta &\approx b(f/v) \end{aligned} \quad (2)$$

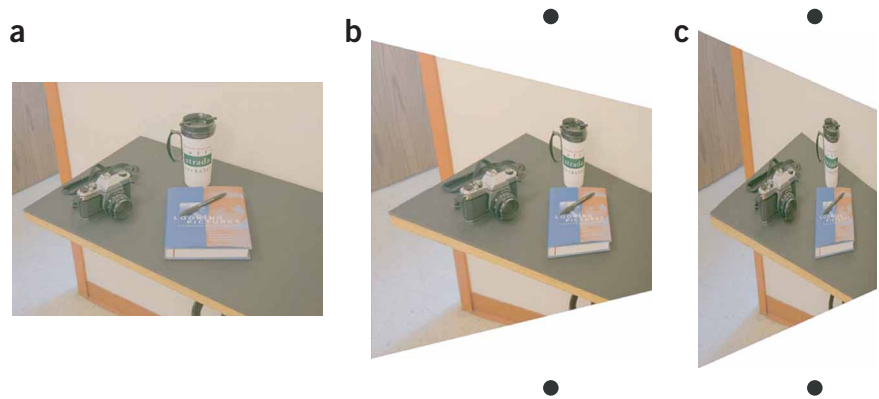
where v is the distance from the eye to the picture surface, f is the eye's focal length and S_{local} is the viewing slant. Thus, the retinal image is foreshortened in the direction of picture-surface slant and scaled according to the viewing distance and focal length.

When a picture is viewed from the CoP, the effects of the two projections (from the object to the projection plane, and from the projection plane to the retina) cancel, so that the dimensions of the retinal image are equal to those that would be created by the actual 3D scene (for example, for a sphere, the retinal aspect ratio α/β remains equal to the depicted ratio A/B at a value of 1). Hence the retinal images

¹Vision Science Program, School of Optometry, and ²Department of Psychology & Helen Wills Neuroscience Center, University of California, Berkeley, California 94720, USA. Correspondence should be addressed to D.V. (dvgss@rit.edu).

Received 28 April; accepted 31 August; published online 18 September 2005; doi:10.1038/nn1553

Figure 1 Demonstration of invariance with oblique viewing. (a) Conventional photograph of an office scene. Focal length was 45 mm for 35-mm film. Hold the photograph at a distance of approximately twice its height and view it with both eyes. Observe the apparent shapes of the cup, book and table. Now rotate it about a vertical axis $\sim 45^\circ$ counterclockwise (bringing the left side nearer) while viewing binocularly. The shapes appear to be relatively undistorted despite the oblique viewing position, which alters the retinal image significantly. View the rotated picture **a** monocularly through a pinhole and the apparent shapes appear rather distorted. (b,c) Photographs of the photograph **a** taken from the approximate locations of each eye when viewing the rotated picture. Hold the page parallel to your forehead at the same distance as before. View **b** with the right eye or **c** with the left eye. The shapes of the objects in the retinal images are about the same as they were with the rotated view of **a**. However, the perceived shapes are more distorted than they were with binocular viewing of the rotated **a**. Now cross-fuse **b** and **c** (direct left eye toward **c** and right eye toward **b**, and then examine the binocularly fused image). Apparent shapes of objects in the fused (and apparently slanted) image are less distorted than when **b** or **c** is viewed separately. The demonstrations are best viewed when projected to large size (so that the CoP distance is greater). Interested readers can obtain the image files and instructions from **Supplementary Figure 8** online.



created by the scene and the picture are identical for an eye at the CoP (**Fig. 2b**). From other viewing positions, the two effects do not cancel, so that the retinal image of the picture differs from the retinal image of the scene (**Fig. 2c,d**). Nonetheless, the perceived aspect ratio (\hat{A}/\hat{B}) remains similar to the depicted ratio (A/B)^{3,7,15,23}. Again, our central question is how such invariance is achieved.

One explanation for invariance is that pictures are typically viewed from small viewing angles (that is, from near the CoP), and thus changes in the retinal image are too small to be noticed^{5,18}. Another explanation states that invariance is a byproduct of the viewer's expectations of familiar shapes (such as faces) or shapes that follow certain geometric rules (right angles, parallel sides, symmetry) and that these expectations allow the viewer to tolerate noticeable distortions^{24,25}. However, the demonstration in **Figure 1a–c** is inconsistent with these explanations.

Another class of explanations proposes compensatory mechanisms that are specific to pictures, based on the claim that an observer at an incorrect viewing position achieves invariance by recovering the true CoP and then reinterpreting the retinal image accordingly. There are two such proposals: the pictorial-compensation and surface-compensation hypotheses.

The pictorial-compensation method uses geometric information present in the picture. Any set of parallel lines in the scene, such as parallel sides of a cube, converges in the projection plane (and therefore in the picture) to produce a vanishing point. The CoP can be computed from three such vanishing points on the picture surface, assuming that the points were generated from orthogonal pairs of parallel lines^{3,4,12,26,27}. Thus, for instance, the CoP could be computed from the vanishing points generated by a cube in the scene. Alternatively, if we assume that the CoP lies on the central surface normal, only two vanishing points are required to determine its location¹⁹. The pictorial-compensation method also requires information about the orientation of the picture surface and its distance from the viewer, because the 3D positions of the vanishing points relative to the observer must be known^{3,4,12}.

The surface-compensation method^{15,25,28} also involves estimating the CoP position but without using the picture's contents. The CoP is assumed to lie on the central surface normal; that is, the picture is assumed to be a normal projection (with the projection plane orthogonal to the optic axis, as in **Fig. 2a,e**). The CoP is determined from a

measurement of the slant of the surface at the middle of the picture, along with an estimate of the distance from this point to the CoP. The distance estimate might be derived from a heuristic based on the size of picture: for example, an assumption that the distance from the CoP to the picture is two times the width of the picture¹⁵.

Once the CoP has been estimated by either the pictorial- or the surface-compensation method, the retinal image is adjusted to account for the distortions resulting from the displacement of the viewing position from the CoP. In principle, pictorial compensation yields geometrically correct compensation in all situations; surface compensation does so only if the picture is a normal projection and if the estimate of the CoP distance to the picture happens to be correct.

This hypothesis does not require estimates of the location of or distance to the CoP. Each point on the picture surface has an orientation with respect to the line of sight from the eye to that point (**Fig. 2e**). That orientation is specified by the slant (S) and the tilt (that is, the direction of the slant)²⁹. The hypothesis states that the visual system estimates the orientation of the picture surface at each point of interest—whether such a point lies in the middle of the picture—and uses these estimates to adjust the dimensions of the image formed on the retina. The adjustment undoes the perspective effects (foreshortening and scaling in the projection to the retina) caused by the slant of the picture surface in the region of interest (equations (2)). For an ovoid, the adjustment yields a perceived width-to-height aspect ratio of

$$\hat{A}/\hat{B} = (\alpha/\beta)/\cos(S_{\text{local}}) \quad (3)$$

In the compensation hypotheses, the retinal image is adjusted only for the portion of the viewing slant caused by the incorrect viewing position: S_{comp} in **Figure 2e**. For an ovoid, the adjustment is:

$$\hat{A}/\hat{B} = (\alpha/\beta)/\cos(S_{\text{comp}}) \quad (4)$$

The two sets of hypotheses—compensation and local slant—make different predictions in many cases. For example, when the eye is at the CoP, the pictorial- and surface-compensation hypotheses claim that no adjustment is made for any region of the picture because $S_{\text{comp}} = 0$. In contrast, the local-slant hypothesis predicts that adjustments will be made in regions away from the picture center because $S_{\text{local}} \neq 0$ at such points. For this reason, the local-slant method does not always yield a geometrically correct estimate, a point we return to later.

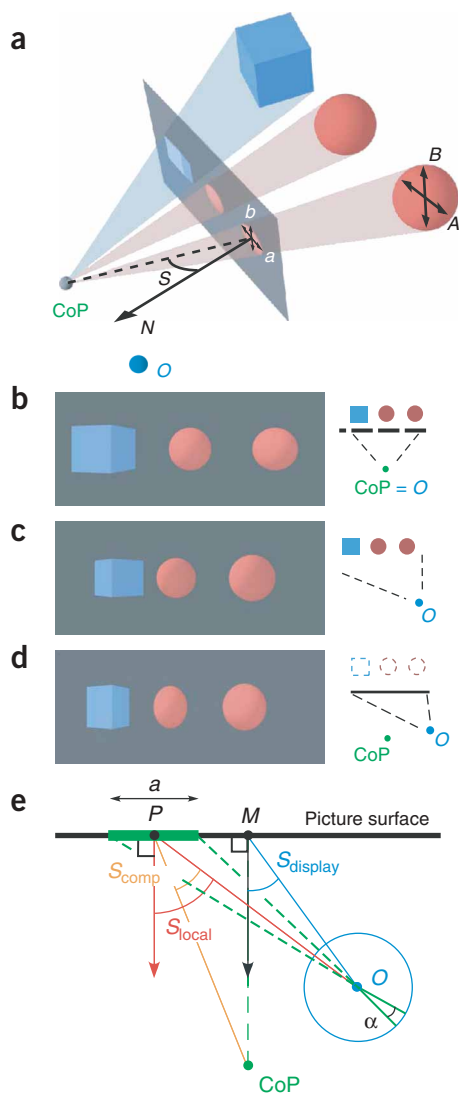


Figure 2 Perspective projection and pictures. (a) Perspective projection of a 3D scene onto a projection plane. The scene is projected toward the CoP, creating a set of light rays called the light field. The picture is the intersection of the light field with the projection plane. The projections of the middle and right spheres create a circle and ellipse, respectively. The scene and picture create the same light field when viewed from the CoP. The orientation of the picture surface at a point can be described by its slant (S , angle between surface normal N and line from CoP to the point) and tilt (direction of slant). In equations (1), the dimensions of the occluding contours of the depicted objects are measured in the tilt direction (A) and orthogonal direction (B)²². a and b are, respectively, the horizontal and vertical dimensions of the image on the projection plane. (b) The image of the original scene for an eye at CoP and the image of the picture for an eye at CoP. The images are identical. The icon on the right is a plan view of the situation. (c) Image of original scene for an eye at O . (d) Image of the picture of the original scene received by an eye at O , and not at CoP. The received image is a double projection: one to create the picture and another to create the retinal image of the picture. (e) Plan view of the projection to the retina. The eye at O views an image on the projection plane. The center of projection is CoP. A picture element with width a is projected forming retinal angle α (equations (2); β not shown here). M is the picture's midpoint; the arrow from M is the central surface normal. P is another point in the picture; the arrow is its surface normal. The surface slants at M and P are $S_{display}$ and S_{local} , respectively.

tion^{15,21,25,26} and local slant⁷—generated different predictions. Invariance was observed with binocular but not monocular viewing. The observed results were consistent with the local-slant hypothesis but not with the pictorial- and surface-compensation hypotheses.

RESULTS

We presented observers with two kinds of objects: slanted planes (Fig. 3a) and ovoids (Fig. 3b). They viewed the stimuli on a fronto-parallel or slanted display (viewing angle = $S_{display}$, Fig. 3c). With the ovoids, observers reported whether the object was too wide or too narrow relative to its height to be a sphere. For judgments to be invariant with changes in viewing angle, the observer must take the viewing obliqueness into account. In principle, the task can be performed with perceived 2D shape; that is, based on judging if the outline of the sphere is a circle on the picture surface. To examine the viewer's ability to segregate the perspective effects caused by oblique viewing from those caused by the picture's contents, we used the slanted-plane task (Fig. 3a). Observers reported whether a rectangular plane rotated about a vertical axis was too wide or too narrow to be a square. For judgments to be invariant, the observer must take into account both the viewing obliqueness and the slant depicted in the picture. Based on trial-by-trial responses, we used a staircase method to determine the aspect ratio (that is, the ratio of horizontal to vertical dimensions) of the ovoid (or plane) that, on average, generated a spherical (or square) percept.

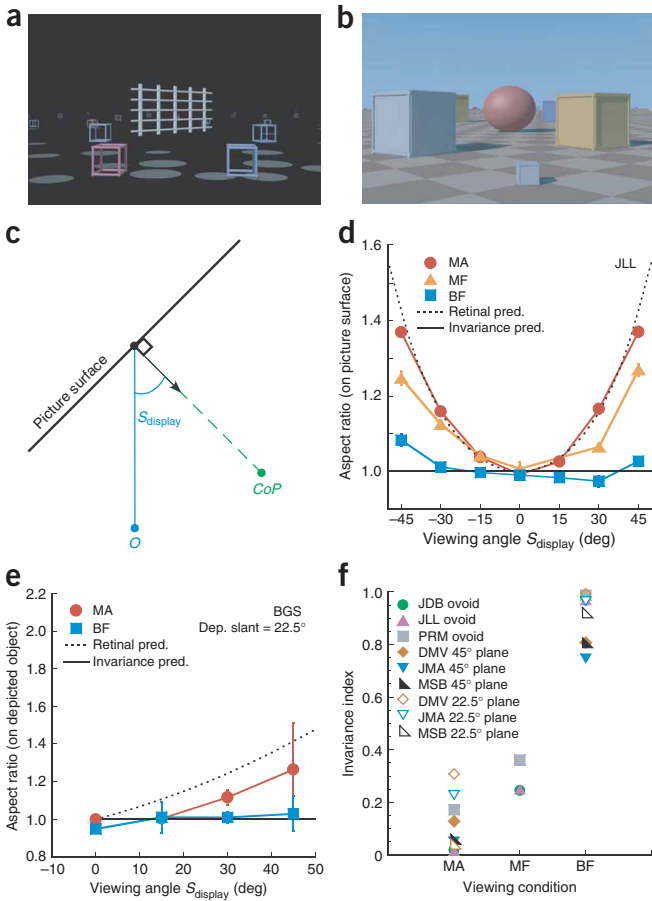
We varied the amount of information available for estimating the slant of the display screen and its distance from the viewer: from least informative to most, the conditions were (i) monocular viewing through an aperture such that the frame of the screen was invisible, (ii) monocular viewing without an aperture such that the frame was visible and (iii) binocular viewing without apertures such that the frame was visible and binocular disparity was available.

Experiment 1: Is perceptual invariance observed?

We first asked whether invariance occurs over a wide range of viewing angles. In this experiment, observers viewed pictures with the display rotated by different amounts ($S_{display}$, Fig. 3c). The pictures were created by normal projection; thus the CoP always lay on the central surface normal.

When pictures are viewed obliquely, perspective in the retinal image is affected by both the slant of the picture surface and the geometry of the scene depicted in the picture. Consider, for example, a picture whose contents are a regularly textured, rectangular plane slanted about the vertical axis (Fig. 3a). The perceived slant of the depicted plane is determined by the foreshortening of individual texture elements and the convergence of horizontal lines in the picture. If one views the picture from a horizontally displaced viewpoint, similar perspective effects are caused by the slant of the picture surface (equations (2)). To perceive the depicted plane properly, the viewer should discount only those perspective effects that result from the obliqueness of the viewpoint and not those caused by the picture's content. The pictorial- and surface-compensation methods segregate the two causes by estimating the position of the CoP and then interpreting the picture from that position. The local-slant method segregates by measuring the slant of the picture surface at the position of interest; two slant cues, both unaffected by the picture's contents, would work for this measurement: binocular disparity and perspective of the picture frame.

We conducted experiments in which we manipulated how pictures were created and viewed. With these manipulations, the three main hypotheses—pictorial compensation^{3,4,12,19,24}, surface compensa-



The results are shown in **Figure 3d–f** and **Supplementary Figures 1** and **2** online. When observers had very little information about the slant of the display (monocular viewing through an aperture), the aspect ratio settings were not invariant over viewing angle, consistent with the observers basing their judgments solely on the shape of the retinal image. When observers had rich slant information (binocular viewing with frame visible), the aspect ratio settings were invariant over viewing angle, particularly when viewing angle was less than $|45^\circ|$; that is, observers based their judgments on a retinal image shape that was first adjusted for incorrect viewing position. When the display was viewed monocularly with the frame visible, the aspect ratios were in between those corresponding to the retinal and invariance conditions, but closer to the retinal. Thus, shapes that looked spherical or square under binocular viewing looked very different under monocular viewing. This disproves the small-distortion^{5,18} and familiar-shape^{24,25} hypotheses (when applied to large slants) because those hypotheses predict no such difference.

The data from the first experiment are summarized in **Figure 3f**. When rich surface-slant information was available, all observers in both tasks showed invariance. When surface-slant information was limited, invariance was not observed in either task: rather, aspect ratios were dictated by the shape of the retinal image. Thus the better the surface-slant information, the greater the invariance; this observation is consistent with earlier work^{23,25}. The relatively small effect of frame visibility (monocular with frame visible condition) is slightly at odds with previous work^{28,30}, a point we discuss later. The results in the slanted-plane task indicate that, when binocular slant information was

Figure 3 Stimuli, predictions and results for the first experiment. (a) Stimulus in slanted plane task. (b) Stimulus in ovoid task with rich pictorial information. (c) Plan view of projection and viewing angles. Projection was frontoparallel and the display screen was rotated about a vertical axis to vary the viewing angle, $S_{display}$. The stimulus was in the display's center. (d) Predictions and results for one observer in ovoid task. Aspect ratio of the ovoid's image on the picture surface is plotted as a function of viewing angle. If invariance occurred, the aspect ratio would lie on the horizontal line ($a/b = 1$) because settings would be circular on the picture surface (invariance predictions). The dotted curve indicates the predicted aspect ratio if the percept were determined by the shape of the retinal image (retinal predictions). In this case, the aspect ratio of the ovoid's image must increase with increases in viewing angle to maintain a circular retinal projection ($\alpha/\beta = 1$). Symbols represent the data: red circles for monocular viewing with picture frame not visible (MA); orange triangles for monocular viewing with frame visible (MF); blue squares for binocular with frame visible (BF). Error bars represent 98% confidence intervals. (e) Predictions and results for slanted-plane task. Data from one observer; depicted slant = 22.5° . The horizontal line (invariance) indicates the predicted aspect ratio if the percept were determined only by the shape of the depicted object in the scene (ratio of 1 on depicted plane). Dotted curve is the predicted aspect ratio if the percept were determined by retinal-image shape only (retinal). (f) Invariance indices for all observers, tasks and viewing conditions. The index is the sum-of-squares error between the data and invariance predictions, normalized by the sum of errors in the invariance and non-invariance predictions. 1 indicates complete invariance and 0 no invariance.

available, observers segregated the perspective effects caused by oblique viewing from those caused by picture contents.

Experiment 2: Does pictorial information underlie invariance?

Having established that binocular viewing yields perceptual invariance despite large changes in viewing angle, we next asked whether pictorial information (that is, the contents of the picture) has a role in the observed invariance. To answer this, we dissociated the geometric information provided by the picture contents from the information provided by surface slant. Specifically, we created pictures in which the projection plane was rotated (S_{proj} in **Figs. 4** and **5a**) and had observers view them from the CoP ($S_{display} = S_{proj}$). An eye at the CoP receives the same pattern of light for all projection angles, (S_{proj}) and so the retinal image does not change no matter what the projection angle is. Because the observer is already at the CoP, pictorial compensation predicts that no adjustments for oblique viewing will occur, and therefore that the settings will follow those predicted from the retinal image alone. The shapes of the objects in the picture, however, change significantly with changes in projection angle (compare **Figs. 4c** and **2b**). As a

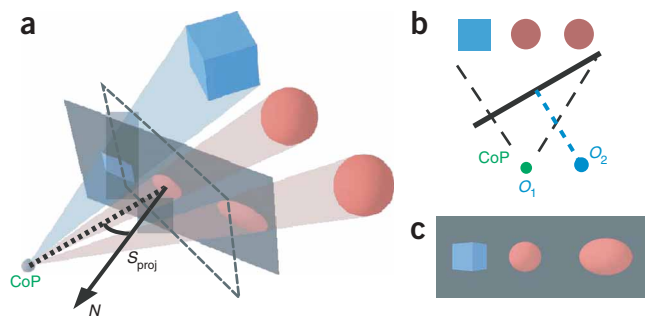
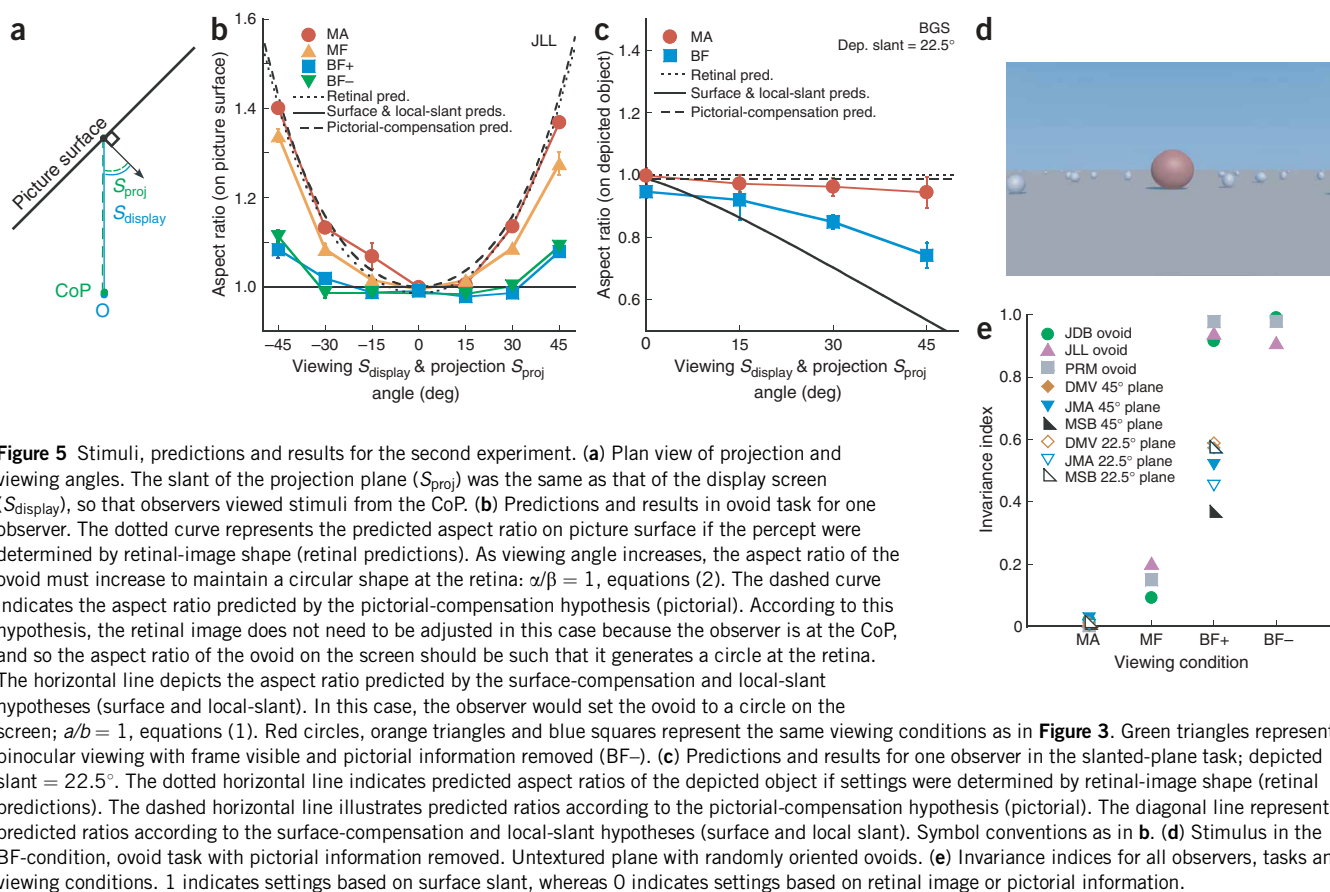


Figure 4 Rotating the projection plane. (a) The same scene as in **Figure 2a**. The projection plane is rotated through angle S_{proj} . (b) Plan view of the situation in **a**. (c) The resulting picture, viewed from point O_2 along the central surface normal rather than from the CoP. Note that it is quite different from the picture of the same scene with normal projection (**Fig. 2b**).



consequence, theories of compensation based on measurement of the surface slant—the surface-compensation and local-slant hypotheses—predict that the pictures will look quite different as S_{proj} changes (because the pictures themselves change).

In this experiment, observers viewed pictures generated with different projection angles from the CoP (such that $S_{display} = S_{proj}$; see **Fig. 5a**). Observers again judged aspect ratios of ovoids and slanted planes. The results for the ovoid task are shown in **Figure 5b** and **Supplementary Figure 3** online. With binocular viewing, the aspect ratios were consistent with the surface-compensation and local-slant predictions, particularly when the viewing angle was less than $|45^\circ|$. This means that a geometrically correct picture viewed binocularly from the CoP looked distorted when the display and projection screens were rotated. Observers behaved as if they were viewing the picture from the picture's central surface normal rather than from the CoP. This observation is clearly inconsistent with the pictorial-compensation hypothesis. With monocular viewing through an aperture, aspect ratios were consistent with the retinal predictions, meaning that a geometrically correct picture viewed from the CoP looked undistorted. In the monocular condition with the frame visible, the data were close to the retinal predictions. The results were very similar in the slanted-plane task (**Fig. 5c**, **Supplementary Fig. 4** online).

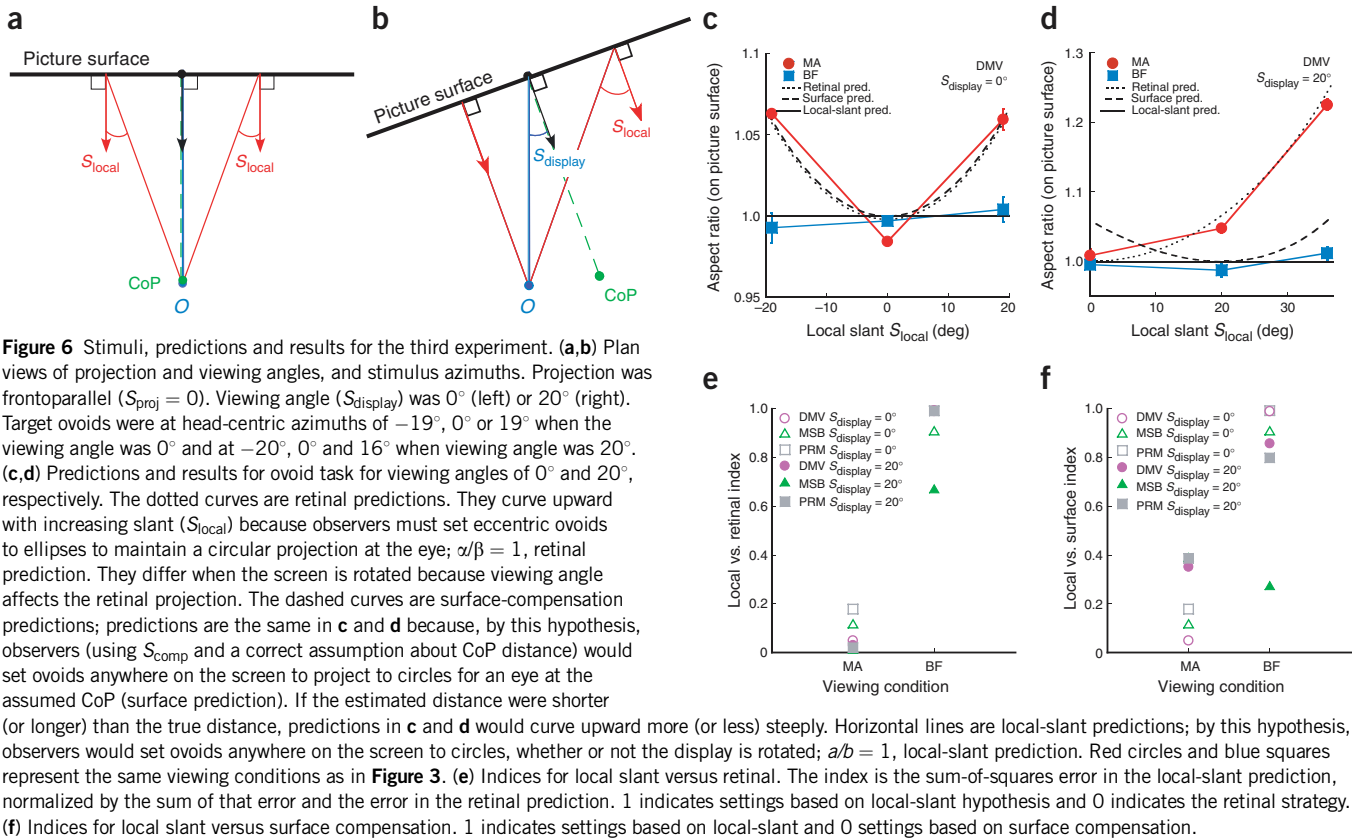
The pictorial-compensation hypothesis requires geometric information in the picture's contents; thus by this hypothesis, more invariance should be observed with geometric information present than with it absent. To test this, we eliminated the information for vanishing points (**Fig. 5d**) and redid the ovoid task. The binocular settings were the same with and without geometric information (**Fig. 5b**, **Supplementary**

Fig. 3), which is inconsistent with the pictorial-compensation hypothesis. We looked still further for evidence of pictorial compensation by creating a condition in which neither the surface-compensation nor the local-slant mechanism would be triggered (**Supplementary Fig. 5** online). Again, the data showed no evidence for pictorial compensation.

The results of the second experiment are summarized in **Figure 5e**. With binocular viewing, observers' aspect ratios were consistent with the surface-compensation and local-slant hypotheses. With monocular viewing, they were consistent with what one would expect if settings were based on the pattern of light striking the eyes (that is, the retinal predictions). In conjunction with the results from the first experiment, this leads us to conclude that adjustment for oblique viewing, when it occurs, is based on surface slant and not on geometric information in the picture. The results thus disprove the pictorial-compensation hypothesis.

Experiment 3: Is the invariance mechanism local?

In the first two experiments, the target objects were presented in the middle of the picture, where $S_{local} = S_{comp}$ and equations (3) and (4) become identical (**Fig. 2e**). For this reason, the results were consistent with both the surface-compensation and local-slant hypotheses. We next tested which of the two provides a better account. We presented target objects at the middle of the display screen (where $S_{local} = S_{comp}$) and toward the edges of the screen (where $S_{local} \neq S_{comp}$; **Fig. 6a,b**). Observers judged the aspect ratios of ovoids on a frontoparallel or slanted display screen (**Fig. 6c–f** and **Supplementary Fig. 6** online). With binocular viewing, aspect ratios were consistent with the local-slant predictions, and with monocular viewing through an aperture they were consistent with retinal predictions. We conclude that



perceptual invariance, when it occurs, is based on measurement of the local slant of the picture surface and adjustments to the perspective effects caused by that slant. As indicated earlier, the local-slant method does not always yield a geometrically correct result, particularly with wide fields of view. We return to this issue later.

Model of adjusting for oblique viewing

An elaboration of the local-slant hypothesis provides an excellent explanation for our data. This model has three parts: (i) estimate the slant and tilt of the picture surface at each point of interest without contamination by 3D cues in the picture’s contents, (ii) adjust the retinal image to, in effect, undo the perspective effects of viewing a slanted picture surface and (iii) interpret the 3D cues in the picture’s contents. Our work concerns the first two parts.

1. Estimating the local slant. Binocular disparity and the perspective of the picture frame are useful cues for estimating the slant of a picture surface without contamination by the picture’s contents. Bayes’ Law prescribes how to weight evidence from various depth cues in order to obtain the most accurate estimate possible^{31–34}. With no immediate consequences of the estimate (that is, payoffs or penalties), the maximum *a posteriori* estimate (MAP) should be the one that is the most probable given the image data and prior information. The MAP estimate for local slant derives from

$$p(S_{local}|i) \propto p(i|S_{local})p(S_{local}) \tag{5}$$

where S_{local} is the local surface slant and i is the input to the eyes. $p(i | S_{local})$ is the likelihood function: the probability of observing various inputs given a particular slant presented to the eyes. $p(S_{local})$ is the prior probability of observing different slants at the eyes based on previous experience. Assuming that disparity and the frame’s

perspective are conditionally independent (that their noises are statistically independent), we can re-write equation (5) as

$$p(S_{local}|i_d, i_f) \propto p(i_d|S_{local})p(i_f|S_{local})p(S_{local}) \tag{6}$$

where d and f refer to the disparity and frame cues, respectively. In our model, the visual system uses the maximum value of $P(S_{local} | i_d, i_f)$ as its estimate of the local slant of the picture surface.

2. Undoing the perspective effects. The MAP estimate from equation (6) is input to equation (3) such that the estimated perspective effects due to oblique viewing are undone. The correctness of this step depends on the accuracy of the slant estimate from equation (6).

3. Interpreting the picture. Once the estimated foreshortening has been undone, the remaining perspective information is used to interpret the picture’s contents. The correctness of the interpretation obviously depends on the accuracy of the preceding steps.

Whenever the standard deviation of either likelihood in equation (6) is much smaller than that of the prior, the MAP estimate is close to the slant presented to the eyes. Whenever the standard deviations are much larger, the estimate approaches the peak of the prior distribution. From the geometry, the prior probability is proportional to $\cos(S_{local})$ —defined from -90° to 90° —because steeply slanted surfaces project to small retinal images³⁴. The half cosine has a peak at $S_{local} = 0$ and standard deviation $\approx 40^\circ$. With binocular viewing, the standard deviation of the disparity likelihood at our distance of 45 cm is $6\text{--}10^\circ$ (ref. 34), which is much smaller than the standard deviation of the prior; thus, the MAP estimate of slant will be close to the presented slant. This would yield nearly complete invariance, as we observed. With monocular viewing through an aperture, the surface slant cannot be estimated reliably; as a result, the standard



© The National Gallery, London

Figure 7 Wide-angle distortions, anamorphic painting and architectural photography. **(a)** Wide-angle photograph of office scene. Lens focal length was 16 mm. CoP distance is two-thirds of the picture height. When viewed binocularly from the CoP, objects near the picture's edges look distorted. Thus wide-angle pictures can appear distorted when viewed correctly. When viewed monocularly from the CoP through a pinhole, objects look much less distorted. It is difficult to view these images from the CoP because that distance is so short. Interested readers should project the images to larger size, thereby increasing the CoP distance. Obtain image files from **Supplementary Figure 8**. **(b)** *The Ambassadors*, by Hans Holbein the Younger. When viewed from straight ahead, one sees a diagonal smear near the bottom. When viewed with one eye from the right and above, the smear is perceived as a skull. **(c)** Conventional photograph of tall buildings (courtesy <http://philip.greenspun.com>). The projection plane is frontoparallel. When viewed binocularly from the CoP (along the central surface normal at a distance of two-thirds the photograph width), the buildings appear to lean toward one another. When viewed monocularly from the CoP through a pinhole, the scene looks more three-dimensional and the buildings lean to a lesser degree. **(d)** Photograph of same scene with projection plane rotated by 21°. View binocularly from straight ahead at same distance as **c**, and the buildings no longer appear to lean. When viewed binocularly from the CoP (21° below the central surface normal), the rotated-projection photograph yields the same retinal images as the conventional photograph **c**, but the perceptual outcome is quite different. When **d** is viewed monocularly through the pinhole from the CoP, the percept is similar to that generated when **c** is viewed from its CoP. Thus, the change in surface slant with CoP viewing of **c** and **d** alters the percept. When the change in slant cannot be detected (monocular viewing through a pinhole), the percepts are more similar because the light fields created by the two photographs are the same. The effects are best tested by projecting **c** and **d** onto a large screen, thereby increasing the CoP distance. Obtain image files from **Supplementary Figure 8**.

deviation of the likelihood is very large and the MAP estimate approaches 0°, the peak of the prior distribution. This would yield no invariance, as we observed.

We observed partial failures of invariance under binocular viewing when the viewing angle was greater than $|45^\circ|$ (**Figs. 3d** and **5b**; **Supplementary Figs. 1** and **3**). At large slants, the disparity gradient becomes large and the ability to fuse the stimulus and estimate its slant is compromised^{35,36}; by our model, the standard deviation of the disparity likelihood increases, so the prior pushes the MAP slant estimate toward 0°, resulting in less invariance.

We found no effect of manipulating the geometric contents of the picture (**Fig. 5b**). This result is consistent with the model because none of

the model's measurements depend on the picture's contents (provided there is sufficient spatial variation to allow disparity measurements).

According to our model, binocular disparity and the perspective of the picture frame can both be used to estimate surface slant. We observed a large effect of disparity but only a small effect of the frame (**Figs. 3d** and **5b**). The latter could have been the consequence of stimulus and task characteristics: observers generally fixated and attended to the target in the center of the display and so may not have picked up the slant information from the frame, which fell in the retinal periphery. This would be expressed as a large standard deviation for the frame likelihood in equation (6), which in turn would yield little effect on the MAP slant estimate.

DISCUSSION

Comparison with previous work

Our results support the local-slant hypothesis and contradict the notion that invariance is achieved by recovering the position of the CoP. Previous studies demonstrated that picture viewing was invariant with respect to viewing obliqueness when surface-slant information was available^{7,14,17,21,23,29}. None showed that the underlying mechanism is local slant, but one researcher came close⁷. One report¹⁷ claimed that invariance for oblique viewing depends greatly on the task. In this study, observers looked at pictures from different viewing angles and reported either the 3D layout of the scene in the picture or the orientation of objects in the scene with respect to the observer. Nearly complete invariance was observed in the layout condition and virtually none in the orientation condition. These results are not incompatible with ours. When observers were asked about the layout of the scene, their judgments should have been made in the coordinates of the depicted scene, and viewing obliqueness should have been taken into account. When asked about orientation relative to the self, the judgment should have been made in observer coordinates where there is no need to take viewing obliqueness into account.

Some previous reports concluded that adjustment for oblique viewing is not based on the slant of the picture surface^{18,20}, which disagrees with our findings. In these studies, observers rated perceptual qualities ("rigidity" and "distortion") of pictures constructed with rotated projection planes. But the authors did not calculate the rigidity or distortion that should have been reported if no adjustment occurred, so one cannot determine whether those data are actually consistent with surface-based adjustment or not.

Does invariance require special pictorial mechanisms?

Investigators have debated whether picture viewing requires special mechanisms^{14–16}. Our data and analysis suggest a new way to evaluate this issue. A person viewing the picture from an arbitrary viewing position must treat separately two perspective effects at the retina: the perspective that is due to oblique viewing and the perspective that is due to the picture's contents. We believe that both of these are manifestations of everyday visual functions and not mechanisms specific to pictures. Adjusting for obliqueness is demonstrated when a person reaches to pick up an object. For instance, consider a book lying in front of the viewer on a desk. The book is slanted relative to the line of sight, so the viewer must first estimate the book's width from the foreshortened retinal image in order to open the hand by the right amount to pick the book up. People are very good at this³⁷; that is, they show shape constancy^{38,39}. Interpreting a picture's contents is a manifestation of inferring 3D layout from the variety of depth cues available in the picture, in a similar way that depth cues from a real object are interpreted.

The mechanism that is special to pictures is that adjustments to oblique viewing seem to occur before the contents of the picture are interpreted, such that the perspective effects are segregated according to their cause. The interpretation of the 3D layout of the picture's contents is thus not contaminated by the perspective distortions caused by oblique viewing.

Distortions with wide-angle pictures

Figure 7a shows another important effect: distortion of perceived object shape in wide-angle pictures, a well-known phenomenon in photography and computer graphics^{6,7,11,13}. The local-slant model predicts such distortions even with geometrically correct pictures viewed binocularly from the CoP.

Photography textbooks recommend choosing a particular lens focal length given the size of the film in order to produce the most useful and realistic photographs. The recommended focal length is usually 40–50% greater than the film width; thus, for example, 35-mm film would require a lens of focal length between 49 and 53 mm^{6,8,9}. What is the basis for this recommendation? Longer focal lengths yield small fields of view when viewed from the CoP and are therefore generally undesirable. But what determines the shortest useful focal length? The textbooks state vaguely that the 40–50% rule creates “a field of view that corresponds to that of normal vision,”⁹ or “the same perspective as the human eye”⁸.

Our analysis offers an explanation for the focal-length recommendation. From the geometry of projection,

$$\theta = 2 \tan^{-1} \left(\frac{w}{2f} \right) \quad (7)$$

where w is film width, f is focal length and θ is the photograph's angular subtense when viewed from the CoP. Assuming that a 5% deviation from the correct aspect ratio is readily detectable⁴⁰, we can determine what field of view yields deviations of this magnitude. From equation (1), the value of S_{local} yielding a deviation of 5% is 18° . To ensure that S_{local} is not larger than 18° , θ must be 36° or smaller. Then for $w = 35$ mm and $\theta = 36^\circ$, $f = 54$ mm, which is quite close to the recommended 49–53 mm. A related recommendation is that a perspective painting, when viewed from the CoP, should subtend no more than $\sim 37^\circ \times 28^\circ$ (refs. 2 and 3), again close to the predicted value.

We suggest that the recommendations for both photography and perspective painting are based on the largest field of view that does not produce perceived distortions due to the local-slant mechanism. One could, in principle, circumvent this problem by creating a display surface for which the local slant is zero everywhere. This would be a hemisphere with the viewer's eyes positioned at the center⁴¹.

The anamorphic effect and architectural photography

The great majority of photographs and paintings are normal projections: the projection plane is perpendicular to a line from the CoP to the picture center (**Figs. 2a,e**; $S_{\text{proj}} = 0$ in **Fig. 4a**). Here we describe two interesting examples in which rotated projections are used ($S_{\text{proj}} \neq 0$).

Anamorphic paintings, like Holbein's *The Ambassadors* (**Fig. 7b**), are images created by large rotations of the projection plane for part of the picture ($S_{\text{proj}} > 60^\circ$)⁴². Viewing Holbein's painting from near the central surface normal, one sees an uninterpretable diagonal smear near the bottom. Most of the painting was created by a normal projection with a CoP on the central surface normal. The smear was created by a rotated projection of a skull with the CoP up and to the right of center. When the viewer moves up and to the right, the smear is seen as a skull.

Our analysis explains why the skull is not perceived as such until the painting is viewed obliquely. For a viewer positioned near the central normal, the smear's retinal image is not at all like the image of a skull and so the skull is not perceived. For a viewer at the skull's CoP, however, the retinal image of the smear is like the image of a skull. But the viewer is also looking oblique to the picture surface, so if complete adjustment for viewing angle occurred, the percept would be more similar to the image on the surface than to the one on the retina, and again the smear could not be interpreted. However, when viewing from the very oblique position of the CoP, the estimate of the local surface slant becomes uncertain because the slant is so great: the best estimate approaches zero (equation (6)), and little if any adjustment for viewing obliqueness occurs (equation (3)). Consequently, a viewer approaching the skull's CoP sees the shape dictated by the retinal image rather than the shape on the surface.

A related phenomenon occurs in architectural photography. In a conventional photograph of tall buildings (**Fig. 7c**), the CoP is on the central normal. Because the camera was pitched upward, the vertical edges of the buildings converge toward the top, creating the disconcerting impression that the buildings are leaning toward one another. Photographers counteract the keystoning effect by rotating the film plane relative to the camera's optic axis^{6,7}. In a photograph of the same scene with the film plane rotated about a horizontal axis (**Fig. 7d**), the CoP is now below the central surface normal. Rotating the film plane is equivalent to rotating the projection plane as we did in the second experiment. In this case, the apparent perspective distortion is undone by the rotation, and the buildings no longer appear to lean inward. Both of these effects—leaning in the conventional photograph and straightening in the rotated photograph—are caused by the viewer perceiving the relationship between the vertical edges on the surface of the picture rather than their relationship in 3D.

METHODS

Six observers participated. Two were authors; the others were unaware of the experimental aims. All but B.G.S. and J.L.L. were experienced psychophysical observers.

Display and viewing parameters. The visual stimuli were generated using ray tracing (POV-Ray) and were displayed on a 21-inch (~ 53 cm) display screen (NEC model FP-2141). Image size was 38.4×24.5 cm when the frame was not occluded. The highest luminances were 0.7 cd/m² for the ovoid stimuli and 2.8 cd/m² for the slanted-plane stimuli. Both were high in contrast, but we cannot summarize the contrast with one number because the stimuli were complex. The display screen was mounted on a rotating and translating platform. We used a sighting device to determine the positions of the centers of rotation of the two eyes relative to the bite bar⁴³. During the experiments, the bite bar was mounted on the same platform as the display screen, so that the viewing distance and angle from the screen were set precisely. The right (viewing) eye was aligned with the center of the display in the monocular conditions, and the cyclopean eye was in the binocular conditions. (All of our predictions in the binocular conditions are based on the images as they would be seen from the cyclopean position. The left and right eyes were slightly displaced from that position, so the predicted settings would be slightly different if the observer used one eye only. With a viewing distance of 45 cm and interocular distance of 6 cm, the change in predicted aspect ratio would be less than 10%.) Images were spatially calibrated for each viewing angle and distance⁴⁴. With this technique, image distortions due to spatial inhomogeneities in the display and prismatic distortion by the faceplate were eliminated.

Viewing distance (the distance from the eye's center of rotation to the center of the phosphor grid of the display screen) was 45 cm for all experiments except the third in which it was 35 cm. Thus, in the first two experiments, the field of view was $45.7^\circ \times 30.5^\circ$ when the display was frontoparallel and the frame was not occluded. In the third experiment, the field of view was $57.0^\circ \times 38.6^\circ$, again when the display was frontoparallel and the frame was not occluded. The CoP distance was 45 cm in the first two experiments and 35 cm in the third. The display was presented at seven viewing angles ($S_{\text{display}} = -45^\circ, -30^\circ, -15^\circ, 0^\circ, 15^\circ, 30^\circ$ and 45°) in the first two experiments and at two angles (0° and 20°) in the third. In each case, the display was rotated about a vertical axis in the center of its phosphor plane, so that the viewing distance did not change. Each viewing angle was presented in a separate experimental session; the sessions were run in random order.

For monocular viewing through an aperture, the room was completely dark and the display was visible only through an 8×6 mm oval aperture placed 1–2 cm from the cornea. The frame of the display was invisible. Observers could not determine the slant of the display screen in this viewing condition. Changes in viewing angle from one session to the next were done behind a curtain so that the observer could not see the change. For monocular viewing without an aperture and binocular viewing without apertures, the room was dimly lit so the entire apparatus, including the display frame, was visible.

Stimulus design. The stimuli in the ovoid task were realistic scenes consisting of a target ovoid lying on a ground plane and 10–20 other objects also lying on the ground plane. Shading was appropriate for surfaces with matte and specular components illuminated by a distant point source and ambient light. In conditions in which we maximized the available pictorial information, the ground plane was textured with square tiles and the background objects were cubes (Fig. 3b); the cubes were rotated randomly about axes perpendicular to the ground plane. There was a random component to the cubes' objective sizes, but their projected sizes were still a cue to distance. The square tiles and cubes created multiple vanishing points on which pictorial compensation relies. In conditions in which we minimized the available pictorial information, the ground plane was untextured and the background cubes were replaced with ovoids, thereby eliminating the vanishing points (Fig. 5c). The size of the target ovoid had a random component so that observers could not perform the task by using only one dimension. The target ovoid appeared in the center of the display screen in the first two experiments, and at three different horizontal positions ($-20^\circ, 0^\circ$ and 16° or 20°) in the third.

The stimuli in the slanted-plane task were realistic scenes consisting of a vertical rectangular plane that lay on a ground plane with 10–20 cubes of random size (Fig. 3c). The cubes were randomly rotated about axes perpendicular to the ground, thereby creating multiple vanishing points. The projected sizes of the cubes were a cue to distance. The vertical plane was textured with a rectangular grid and had a depicted slant of 22.5° or 45° . The plane always appeared in the center of the screen. Its size, number of cells and grid thickness had random components so that observers could not perform the task by using only one dimension. A control experiment validated the usefulness of this task (Supplementary Fig. 7 online).

Procedure. Observers initiated each trial with a button press. The stimulus appeared for either 2 s (ovoid) or 1 s (slanted plane). The stimulus was then extinguished and the observer made a two-alternative, forced-choice response indicating whether the stimulus was wider or narrower than a sphere or square. The next stimulus

appeared 1 s after the response. No instructions were given about where to look during stimulus presentations, but observers reported that they looked at the target object. The aspect ratio of the stimulus was varied according to an adaptive one-up, one-down staircase until eight reversals occurred. In the ovoid task, the ratio was varied in the coordinates of the display screen. The observers were given no instructions in this task as to which coordinate system they should base their judgments on; none reported difficulty in doing the task, and all produced repeatable settings. In the plane task, the staircase varied aspect ratio in the coordinates of the depicted plane. The observers were instructed in this task to base their judgments on the dimensions of the depicted 3D object.

A cumulative Gaussian was fit to the data from each staircase using a maximum-likelihood criterion⁴⁵. The estimate of the aspect ratio that on average looked most spherical or most square-like was the mean of the Gaussian. Error bars were the 98% confidence intervals⁴⁶.

Note: Supplementary information is available on the Nature Neuroscience website

ACKNOWLEDGMENTS

We thank M. Landy, J. Hillis, A. Welchman, R. Fleming, S. Gepshtein and M. Ernst for comments on an earlier draft, and R. Bartholomew for technical assistance. This work was supported by NIH research grant R01-EY014194 (M.S.B.), NIH post-doctoral fellowship F32 EY14514 (D.V.) and by DOE Computational Sciences Graduate Fellowship DE-FG02-97ER25308 (A.R.G.).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/natureneuroscience/>
Reprints and permissions information is available online at <http://ngp.nature.com/reprintsandpermissions/>

1. da Vinci, L. *The Literary Works of Leonardo da Vinci* (ed. Richter, J.P.) (Phaidon, London, 1970).
2. Olmer, P. *Tracés Pratiques* (Plon, Paris, 1949).
3. Kubovy, M. *The Psychology of Perspective and Renaissance Art* (Cambridge Univ. Press, New York, 1986).
4. La Gournerie, J.D. *Traité de Perspective Linéaire Contenant les Tracés pour les Tableaux, Plans et Courbes, les Bas-reliefs et les Décorations Théâtrales, avec une Théorie des Effets de Perspective* (Dalmont et Dunod, Paris, 1859).
5. Gombrich, E.H. *Art and Illusion* (Princeton Univ. Press, Princeton, New Jersey, 1960).
6. Kingslake, R. *Lenses in Photography: The Practical Guide to Optics for Photographers* (Case-Hoyt, Garden City, New York, 1951).
7. Pirenne, M.H. *Optics, Painting and Photography* (Cambridge Univ. Press, Cambridge, UK, 1970).
8. Alesse, C. *Basic 35mm Photo Guide: For Beginning Photographers* (Amherst Media, New York, 1989).
9. Giancoli, D.C. *Physics: Principles with Applications* (Prentice Hall, Englewood Cliffs, New Jersey, USA, 2000).
10. Meister, R. The iso-deformation of images and the criterion for delimitation of the usable areas in cine-auditoriums. *J. Soc. Motion Pict. Television Eng.* **75**, 179–182 (1966).
11. Zorin, D. & Barr, A.H. Correction of geometric perceptual distortions in pictures. in *Proceedings of SIGGRAPH 1995* Vol. 14 (ed. Cook, R.) 257–264 (ACM SIGGRAPH/Addison-Wesley, Boston, 1995).
12. Caprile, B. & Torre, V. Using vanishing points for camera calibration. *Int. J. Comput. Vis.* **4**, 127–140 (1990).
13. Agrawala, M., Zorin, D. & Munzner, T. Artistic multiprojection rendering. *Proceedings of the 11th Eurographics Rendering Workshop* **11**, 125–136 (2000).
14. Hagen, M.A. Influence of picture surface and station point on the ability to compensate for oblique view in pictorial perception. *Dev. Psychol.* **12**, 57–63 (1976).
15. Rosinski, R.R. & Farber, J. Compensation for viewing point in the perception of pictured space. in *The Perception of Pictures* (ed. Hagen, M.A.) 137–176 (Academic, New York, 1980).
16. Rogers, S.J. Perceiving pictorial space. in *Perception of Space and Motion* (eds. Epstein, W. & Rogers, S.J.) 119–163 (Academic, San Diego, 1995).
17. Goldstein, E.B. Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *J. Exp. Psychol. Hum. Percept. Perform.* **13**, 256–266 (1987).
18. Cutting, J.E. Rigidity in cinema seen from the front row, side aisle. *J. Exp. Psychol. Hum. Percept. Perform.* **13**, 323–334 (1987).
19. Sedgwick, H.A. The effects of viewpoint on the virtual space of pictures. in *Pictorial Communication in Virtual and Real Environments* (eds. Ellis, S.R., Kaiser, M.K. & Grunwald, A.C.) 460–479 (Taylor & Francis, London, 1991).

20. Gershun, A. The light field. *J. Math. Phys.* **23**, 51–151 (1939).
21. Gibson, J.J. *The Perception of the Visual World*. Houghton-Mifflin, Boston, (1950).
22. Gårding, J. Shape from texture for smooth curved surfaces in perspective projection. *J. Math. Imaging Vis.* **2**, 329–352 (1992).
23. Rosinski, R.R., Mulholland, T., Degelman, D. & Farber, J. Picture perception: An analysis of visual compensation. *Percept. Psychophys.* **28**, 521–526 (1980).
24. Busey, T.A., Brady, N.P. & Cutting, J.E. Compensation is unnecessary for the perception of faces in slanted pictures. *Percept. Psychophys.* **48**, 1–11 (1990).
25. Perkins, D.N. Compensating for distortion in viewing pictures obliquely. *Percept. Psychophys.* **14**, 13–18 (1973).
26. Adams, K.R. Perspective and the viewpoint. *Leonardo* **5**, 209–217 (1972).
27. Greene, R. Determining the preferred viewpoint in linear perspective. *Leonardo* **16**, 97–102 (1983).
28. Wallach, H. & Marshall, F.J. Shape constancy in pictorial representation. *Percept. Psychophys.* **39**, 233–235 (1986).
29. Stevens, K.A. Slant-tilt: The visual encoding of surface orientation. *Biol. Cybern.* **46**, 183–195 (1983).
30. Koenderink, J.J. & van Doorn, A.J. Pictorial space. in *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space* (eds. Hecht, H., Schwartz, R. & Atherton) 239–299 (MIT Press, Cambridge, Massachusetts, USA, 2003).
31. Kersten, D., Mamassian, P. & Yuille, A. Object perception as Bayesian Inference. *Annu. Rev. Psychol.* **55**, 271–304 (2004).
32. Landy, M.S., Maloney, L.T., Johnston, E.B. & Young, M. Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Res.* **35**, 389–412 (1995).
33. Knill, D.C. & Saunders, J.A. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res.* **43**, 2539–2558 (2003).
34. Hillis, J.M., Watt, S.J., Landy, M.S. & Banks, M.S. Slant from texture and disparity cues: optimal cue combination. *J. Vis.* **4**, 967–992 (2004).
35. Burt, P. & Julesz, B. A disparity gradient limit for binocular fusion. *Science* **208**, 615–617 (1980).
36. Banks, M.S., Gepshtein, S. & Landy, M.S. Why is spatial stereoresolution so low? *J. Neurosci.* **24**, 2077–2089 (2004).
37. Mamassian, P. Prehension of objects oriented in three-dimensional space. *Exp. Brain Res.* **114**, 235–245 (1997).
38. Thouless, R.H. Phenomenal regression to the “real” object. *Br. J. Psychol.* **21**, 339–359 (1931).
39. Epstein, W.P. & Park, J.N. Shape constancy: Functional relationships and theoretical formulations. *Psychol. Bull.* **60**, 265–288 (1963).
40. Regan, D. & Hamstra, S.J. Shape discrimination and the judgment of perfect symmetry: dissociation of shape from size. *Vision Res.* **32**, 1845–1864 (1992).
41. Evans, F. & Narayanan, A. Immersive data visualization with the VisionDome. *SPIE Visual Data Exploration and Analysis VII* (SPIE, San Jose, California, USA, 2000).
42. Topper, D. On anamorphosis: Setting some things straight. *Leonardo* **33**, 115–124 (2000).
43. Hillis, J.M. & Banks, M.S. Are corresponding points fixed? *Vision Res.* **41**, 2457–2473 (2001).
44. Backus, B.T., Banks, M.S., van Ee, R. & Crowell, J.A. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Res.* **39**, 1143–1170 (1999).
45. Wichmann, F.A. & Hill, N.J. The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception and psychophysics. Percept. Psychophys.* **63**, 1293–1313 (2001).
46. Wichmann, F.A. & Hill, N.J. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept. Psychophys.* **63**, 1314–1329 (2001).