Where should you sit to watch a movie?

Martin S. Banks[*a,b], Heather F. Rose[c], Dhanraj Vishwanath[a], & Ahna R. Girshick[a]
[a]Vision Science Program, University of California, Berkeley, CA 94720-2020, USA
[b]Dept. of Psychology & Wills Neuroscience Center, University of California, Berkeley, CA 94720, USA
[c]Bioengineering Program, University of California, Berkeley, CA 94720-1762, USA

[*] marty@john.berkeley.edu; phone 510-642-9341

ABSTRACT

A picture viewed from its center of projection (CoP) generates the same retinal image as the original scene. When a picture is viewed from other locations, the retinal image specifies a different layout and shapes, but we normally do not notice the changes. The mechanism underlying this is unknown. We studied the perceived shapes of pictured ovoids and planes while varying viewing angle and the angle by which the pictures were projected. We also varied the viewer's information about the orientation of the picture surface. Viewers compensated nearly veridically for oblique viewing when binocular information for surface orientation was available. In so doing, they used an estimate of local surface orientation and not prior information for object shape nor geometric information in the picture. We present a model that explains invariance for incorrect viewing positions, and other phenomena like perceived distortions with wide fields of view.

1. INTRODUCTION

Pictures have widespread usage because in the convenient format of a 2D surface they allow viewers to perceive the 3D layout and shapes of objects in a scene. Pictures would not be very useful if the observer's eye had to be positioned at the CoP. Imagine, for example, there being only one seat in the theater that allowed an acceptable impression of a motion picture. Painters (da Vinci, 1970; Kubovy, 1986; Gombrich, 1960), photographers (Pirenne, 1970; Kingslake, 1951), cinematographers (Meister, 1966), computer scientists (Zorin & Barr, 1995; Caprile & Torre, 1990), and vision scientists (Hagen, 1976; Rosinki & Farber, 1980; Goldstein, 1987; Cutting, 1987) have for years wondered how perceptual invariance across viewing position is achieved with pictures; they have also developed rules of thumb for minimizing apparent distortions. Our work examines the means for achieving invariance and the causes of perceived distortions.

1.1 Perspective projection

The underlying principle for creating pictures is *perspective projection*. In Figure 1a light rays from points in the scene are "projected" toward the CoP, creating the *light field* (Gershun, 1939) or optic array (Gibson, 1950). The intersection of the light field with the projection plane is the *picture*. The scene and objects that generated the picture are the *depicted scene* and *depicted objects*. The picture reproduces the light field of the original scene for a camera or eye at the CoP. Translations and rotations of the projection plane yield different pictures, but the view from the CoP, which is the same position in each case, remains the same. A *normal projection* is when the projection plane is perpendicular to the line from the CoP to the middle of the picture.

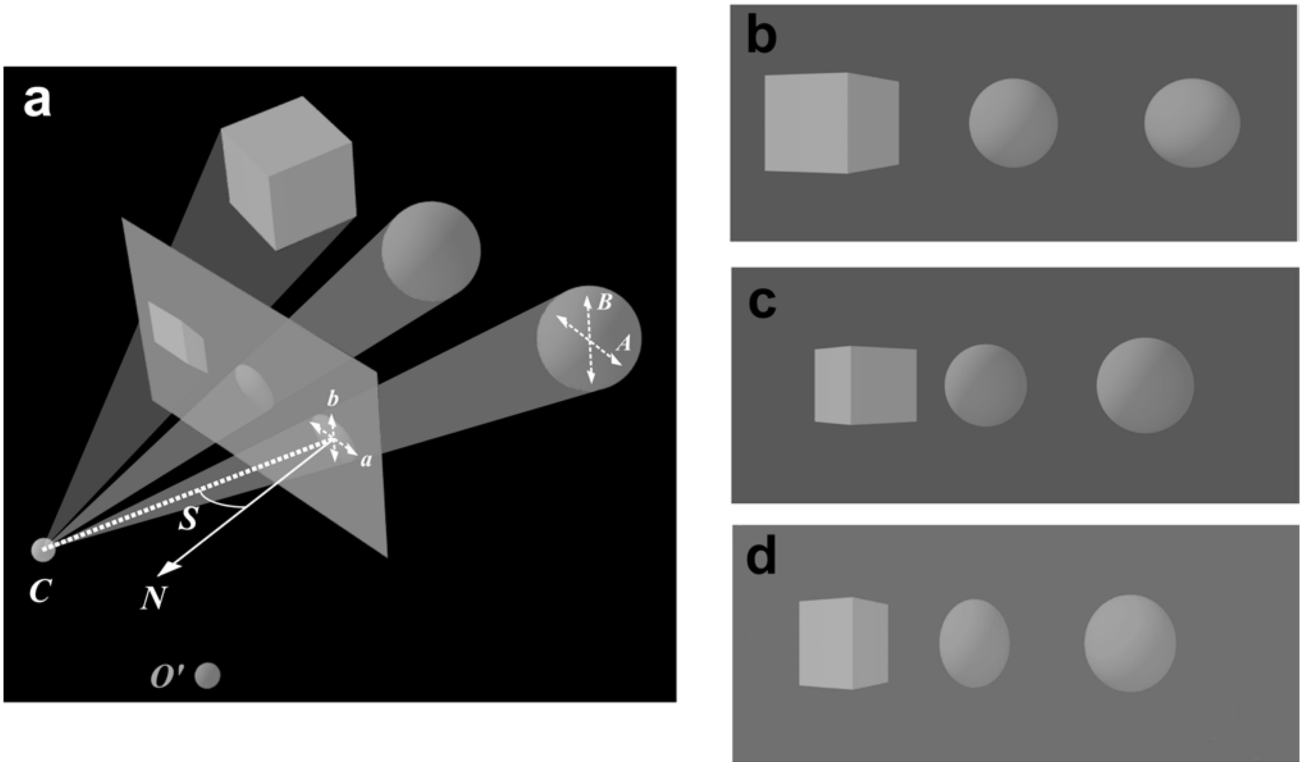

Figure 1. Perspective projection and pictures. a) The geometry of perspective projection. A 3D scene is on the right, projection plane in the middle, and center of projection (CoP) on the left. The scene is projected toward the CoP creating a set of light rays called the *light field*. The projections of the occluding contours of the spheres are cones with apices at the CoP. The projection of the occluding contours of the cube is a six-sided object with an apex at the CoP. Contours in the picture are determined by the intersections of the cones and six-sided object with the projection plane. The cone associated with the central sphere is centered on a line perpendicular to the projection plane and through the CoP, so its intersection with the projection plane is a circle. The cone associated with the sphere on the right pierces the projection plane non-perpendicularly, so the intersection is an ellipse. The scene and picture create the same light field when they are viewed from the CoP. Translations and rotations of the projection plane do not alter the light field created by the projection when it is viewed from the same position (the CoP). The orientation of the picture surface at a point can be described by slant and tilt at that point. Slant S is the angle between the surface normal η and the line from the CoP to the point. Tilt is the direction of slant: specifically, the direction from the standpoint of the CoP in which distance is changing most rapidly around the point. In Eqns. 1 we adopt the convention that the dimensions of the occluding contours of the depicted objects are measured in the tilt direction (A) and the orthogonal direction (B). p and d are the distances along the line to the point of interest from the CoP to the projection plane and the object in the scene, respectively. a and b are the dimensions of the image on the projection plane. Because of this convention, the foreshortening occurs in a and not b. b) Picture of the original scene in a for a camera placed at CoP. This is identical to a picture of the projection plane in a as long as the camera is at CoP. c) Picture of the original scene for a camera placed at O. It is not a picture of the projection plane in a. d) Picture of the projection plane in a for a camera at O. It is therefore a double projection: one projection to create the picture on the projection plane and another to create the picture of the picture. c and d are quite different pictures.

3D orientation at a point on the picture surface is given by the slant and tilt of the surface at that point. To close approximation, the effects of projection onto a plane can be described by two effects: projected size and projected shape (Gårding, 1992). The dimensions of the occluding contours of the depicted object when measured in the tilt direction and in the orthogonal direction are, respectively, A and B. For the right sphere in Figure 1a, tilt is horizontal, so A corresponds to the diameter measured horizontally and B to the vertical diameter. The dimensions of the image at the projection plane are:

$$a = A(p/d)/\cos(S)$$
$$b = B(p/d) \qquad (1)$$

where d is the sphere's distance from the CoP, p is the distance of the projection plane from the CoP along a line toward the sphere, and S is the slant of the plane at the point of interest. Eqns. 1 become exact as the size of the region in the picture approaches zero. Eqns. 1 show that increasing the distance from the projection plane to the CoP yields a larger image; this applies equally to a and b. Increasing the slant of the projection plane in the region of interest yields greater foreshortening (actually lengthening) in the tilt direction; this effect, expressed by $1/\cos(S)$, applies to a and not b. The aspect ratio a/b of a small element in the picture is $A/[B\cos(S)]$, which for a sphere is $1/\cos(S)$.

Figs. 1b-d show the similarities and differences between viewing a scene and viewing pictures of the same scene. Figure 1b is the view seen by a camera placed at CoP in Figure 1a. The view is the same for the scene and the picture of the scene. Figure 1c is the view of the scene when the camera is at O. Figure 1d is the view of the picture of that scene created when the camera was at CoP (Fig. 1b), but with the picture viewed from O. It is a double projection: the first projection occurs in the creation of the picture, and the second with the oblique viewing of the picture. The resulting pictures in Figs. 1c,d are now quite different. The means by which the visual system recovers the 3D scene from oblique views of pictures, like Figure 1d, is the central question examined here.

1.2 Explanations of perceptual invariance

Several explanations have been offered for the apparent invariance of perceived layout and shape in pictures with changes in viewing position.

Small-distortion hypothesis. Proponents of this hypothesis argue that the slant at which pictures are viewed is usually small, and consequently the distortions of the retinal image are too small to be noticed (Gombrich, 1960; Cutting, 1987). Presumably, they would argue that the distortions would be seen if the picture were viewed at a large slant. Consider viewing a circle from an oblique position such that the slant of the picture where the circle lies is S. By Eqns. 1, the aspect ratio of the light field approaching the eye is $\cos(S)$, which is ~1 for slants less than 20 deg. So for small slants, the change in the projection of the picture element from its correct value is small and arguably not noticeable.

Familiar-shape hypothesis. This hypothesis claims that invariance is a byproduct of the viewer's expectations with known shapes (Perkins, 1973). For example, if the retinal image is similar to the image that would be created by a cube, prior expectations force the percept to that of a cube. Invariance thus comes from the viewer's experience with objects whose shapes are familiar (faces) or usually follow certain rules (right angles, parallel sides, symmetry).

Compensation hypotheses. These hypotheses claim that invariance is the consequence of altering or re-interpreting the retinal image by recovering the position of the CoP from information in the picture or from the orientation of the picture surface. There are two versions: pictorial compensation and global surface compensation.

According to the pictorial hypothesis, the CoP is recovered from the locations of vanishing points in the light field: i.e., from geometric information in the picture's contents. The locations of three orthogonal vanishing points are sufficient (Kubovy, 1986; Caprile & Torre, 1990). Alternatively, two orthogonal vanishing points plus the assumption that the CoP lies on the surface normal from the center of the picture can be used (Sedgwick, 1991). (The 3D positions of the vanishing points must be known, so this method also requires knowing the slant and distance of the picture surface.)

The global surface hypothesis (Rosinski & Farber, 1980; Wallach & Marshall, 1986) also recovers the position of the CoP, but not by using the picture's contents. The slant at the middle of the picture's surface is measured from cues such as binocular disparity. The direction of the CoP from the surface can then be determined from the slant measurement, but its distance must be inferred from heuristics such as assuming that the picture should subtend a particular field of view and then calculating at what distance it would (Rosinski & Farber, 1980).

Once the position of the CoP is estimated from either method, the observer could compensate by the means formalized in Eqns. 2 and 3. The part of the surface slant at a point P that requires compensation for foreshortening is

$$\boldsymbol{S_{comp} = \cot^{-1}\{\frac{d_O}{d_C}[\tan(S_{OM}+\gamma_{OP})-\tan(S_{OM})]\} - \cot^{-1}[\tan(S_{OM}+\gamma_{OP})]} \,. \qquad (2)$$

The measurements a and b in Eqns. 1 are then adjusted into

$$\begin{aligned} \hat{a} &= a/\cos(S_{comp}) \\ \hat{b} &= b \end{aligned} \qquad (3)$$

and $\hat{a}$ and $\hat{b}$ are used to interpret elements in the picture. The pictorial-compensation method yields the correct compensation for oblique viewing. The global surface method does too if the CoP lies on the center normal and the distance d_C is inferred accurately.

Local surface hypothesis. This hypothesis is fundamentally different from the compensation hypotheses. According to the local hypothesis, the visual system estimates the slant and tilt at each point of interest, whether in the middle of the picture or not, and then undoes the foreshortening at each point. The adjustment or re-interpretation of the retinal image, expressed in Eqns. 3 with S_{comp} equal to the local slant, occurs for any region on a picture whose slant differs from 0. The local surface hypothesis does not require estimates of d_O and d_C. Importantly, the retinal-image adjustments do not yield the geometrically correct compensation for oblique viewing, a point we return to later. Photographic demonstrations by Pirenne (1970) exemplify the local surface mechanism, but he did not articulate a theory.

1.3 The need to segregate slant cues in picture viewing

When viewing pictures 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 plane slanted about the vertical axis (Fig. 4c). The foreshortening of individual texture elements and the convergence of horizontal lines are cues to the plane's slant. If one views the picture from a horizontally displaced viewpoint, the same foreshortening and convergence effects are caused by the slant of the picture surface. To perceive the depicted plane properly, the viewer should compensate or adjust for only the perspective effects caused by the viewing obliqueness and not for those caused by the picture's content. The compensation and local surface hypotheses state how that segregation could occur. Pictorial compensation takes advantage of the fact that the two causes of perspective effects are separable if there is sufficient geometric information in the picture's contents. The global and local surface hypotheses use the measured slant of the picture surface; two slant cues—binocular disparity and perspective of the picture frame—would work because they signal the slant of the picture's surface and not the slant of its contents.

2. METHODS & RESULTS

We investigated which hypothesis is the best account for the perceptual invariance associated with oblique viewing of pictures. We also developed a quantitative model of the process; the model takes into account measurement uncertainty and prior information.

The experimental tasks allowed us to distinguish the above-mentioned hypotheses. There were two kinds of objects: ovoids (Fig. 2b) and slanted planes (not shown). With the ovoids, observers reported whether the object was too wide relative to its height (or vice versa) to be a sphere. The ovoid task can be performed with perceived 2D shape. To examine the ability to segregate perspective effects caused by oblique viewing from perspective effects due to the picture's contents, we used the slanted-plane task. Observers reported whether a rectangular plane rotated about a vertical axis (tilt = 0) was too wide or too narrow to be a square in the depicted scene. To perform this task veridically, the observer must take into account both the viewing obliqueness and the slant depicted in the picture.

Based on trial-by-trial responses, a staircase adjusted the aspect ratio of the objects to determine the ratio that on average looked spherical or square. Because observers had to respond "too wide" or "too narrow" on each trial, the task eliminates the influence of prior expectations, which would force a wide range of aspect ratios to appear spherical or square (Perkins, 1973).

The display could be rotated about a vertical axis through the center of its front surface (Fig. 2a). Because the rotations were as large as |45| deg, we could determine whether adjustments for obliqueness occur at large angles and thereby test the small-distortion hypothesis (Cutting, 1987). We varied the amount of information available for estimating the slant and distance of the display screen: from least informative to most: a) monocular viewing through an aperture such that the frame of the screen was invisible, b) monocular viewing without an aperture such that the frame was visible, and c) binocular viewing without apertures such that the frame was visible and binocular disparity was available.

2.1 Experiment 1: How accurate are adjustments for oblique viewing?

We first asked whether quantitatively correct adjustments occur when viewing at large angles. Observers viewed normal projections of the simulated scenes with the display rotated by different amounts. The CoP was always on the central surface normal of the display (Fig. 2a).

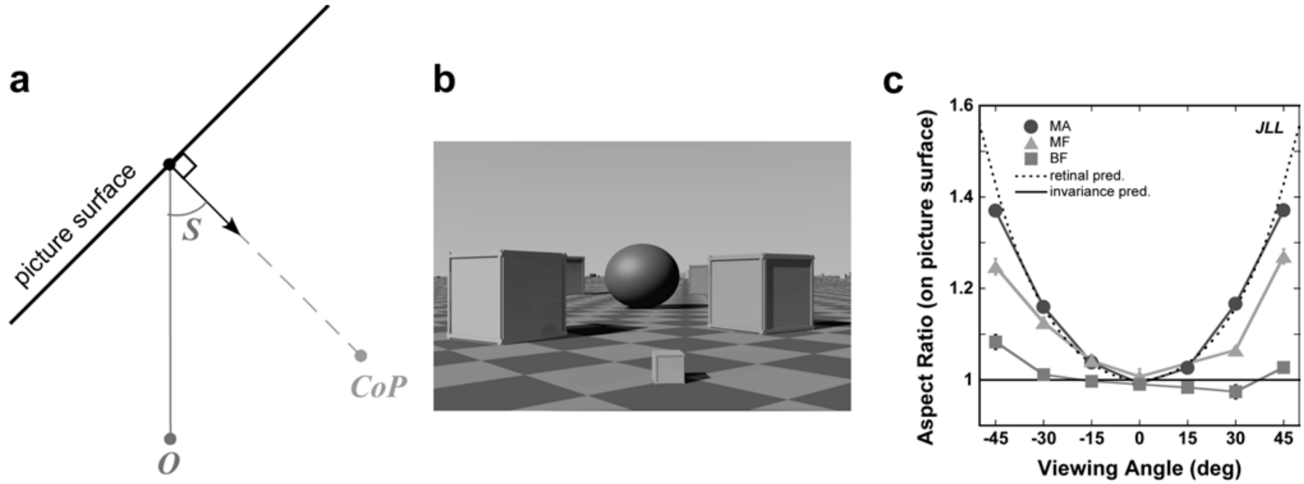

Figure 2. Stimuli, predictions, and results for the first experiment. a) Plan view of the projection and viewing angles in this experiment. The projection was always frontoparallel (i.e., normal) and the display screen was rotated about a vertical axis through its face to vary the viewing angle. Because the stimulus was in the center of the display, the viewing angle was equal to the slant, *S*. b) An example of the stimulus in the ovoid task when the pictorial information was rich; the ground plane consisted of square tiles and many randomly oriented cubes were placed on the ground plane. c) Predictions and results for the ovoid task. The dotted curve is the predicted aspect ratios of the settings in screen coordinates if the settings were determined by the shape of the retinal image (*retinal predictions*). The settings would be those that generate a conical light field (thereby generating the retinal image associated with a circle). Such settings would fall on the dotted curve because, as the viewing angle increases, the aspect ratio of the ovoid must increase in order to maintain a conical projection toward the eye (Eqns. 1). The horizontal line is the predicted settings if invariance occurred (*invariance predictions*). The settings would be circular on the screen and hence would always have an aspect ratio of 1 (in screen coordinates). The symbols represent the results for one observer. Settings with monocular viewing through an aperture (frame not visible; MA) are represented by the circles. Settings with monocular viewing without an aperture (frame visible; MF) are represented by triangles. Binocular settings without apertures (frame visible; BF) are represented by squares. Error bars represent 98% confidence intervals.

The predictions and results for the ovoid task are shown in Figure 1c. With monocular aperture viewing, where the observer has very little information about the slant of the display screen, ovoid settings followed the retinal predictions (no adjustment for oblique viewing). With binocular viewing, where the observer has rich slant information, settings followed the invariance predictions (veridical adjustment), particularly when the viewing angle was less than |45 deg|. These data replicate earlier work in which no adjustment was observed with monocular viewing and nearly veridical adjustment with binocular viewing (Rosinski, Mulholland, Degelman, & Farber, 1980). With monocular viewing and the frame visible, the data fell in between the retinal and invariance predictions, but much closer to the retinal. Thus, frame information had only a small effect, which is slightly at odds with previous work (Wallach & Marshall, 1986; Koenderink & van Doorn, 2003); we reconsider this later.

To examine the ability to segregate perspective effects caused by oblique viewing from perspective effects due to picture contents, we used the slanted-plane task. The plane was a small rectangular grid rotated about the vertical axis. Observers judged the aspect ratio of the plane, but while viewing the display from various angles. As with the ovoid task, the results were consistent with the retinal predictions when information for surface slant was unavailable (monocular viewing through an aperture) and consistent with the invariance predictions when slant information was available (binocular with frame visible). Thus, people can segregate the perspective effects caused by oblique viewing from those caused by picture contents when binocular slant information is available.

In summary, we found that shapes that look spherical or square under binocular viewing look very different under monocular viewing. This finding disconfirms the small-distortion (Gombrich, 1960; Cutting, 1987) and familiar-shape (Perkins, 1973) hypotheses (when applied to large slants) because they predict no such difference.

2.2 Experiment 2: Does pictorial or surface information govern adjustment for oblique viewing?

We next sought to determine whether pictorial compensation or the surface hypotheses provide the best account of perceptual invariance. They all predicted invariance in the first experiment, so we needed another manipulation to tease them apart. We achieved this by using rotated projections, a technique used in architectural photography (Kingslake, 1951). As implied by Fig. 1, rotation of the projection plane does not affect the image seen at the CoP: the same light field is experienced whether one views a frontoparallel projection from the picture's central normal or a rotated projection with the display screen rotated by the same amount. From the standpoint of the pictorial-compensation hypothesis, changes in the projection angle coupled with changes in viewing angle will not affect percepts because the observer is always at the CoP. In contrast, the global and local surface hypotheses predict large differences because the slant of the picture surface has been altered.

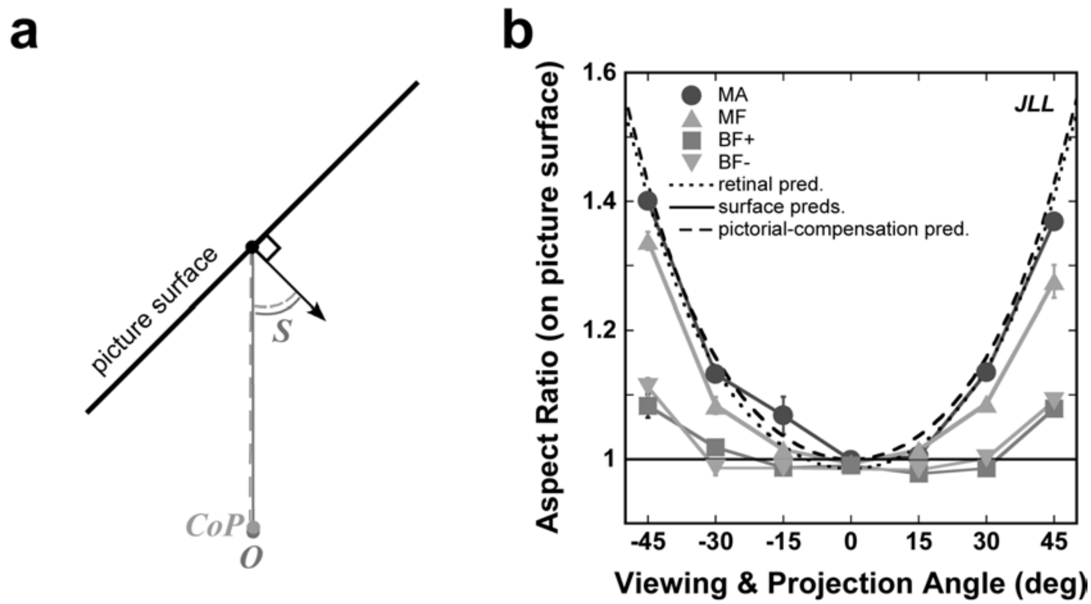

Figure 3. Stimuli, predictions, and results for the second experiment. a) Plan view of the projection and viewing angle in this experiment. The angles of the projection plane (ρ) and of the display screen (*S*) were always the same, so observers viewed the stimuli from the CoP (or in binocular conditions, with the CoP at the cyclopean eye). b) Predictions and results for the ovoid task. The dotted curve is the predicted aspect ratios of the settings in screen coordinates if the settings were determined by the shape of the retinal image (*retinal predictions*). The settings would be those that generate a conical light field (thereby generating the retinal image associated with a circle; Fig. 1). Such settings would fall on the dotted curve because, as the viewing angle increases, the aspect ratio of the ovoid must increase in order to maintain a conical projection toward the eye (Eqns. 1). The predictions of the pictorial-compensation hypothesis are represented by the dashed curve (*pictorial predictions*). The observer was at the CoP, so if adjustment for oblique viewing were based on geometric information in the picture, he/she would set the ovoid such that it generated the retinal image associated with a circle. The data would fall on the dashed curve because, as viewing angle increases, the aspect ratio of the ovoid on the screen must increase in order to maintain a conical projection toward the eye. The predictions for the global and local surface hypotheses are represented by the horizontal line (*surface predictions*). According to those hypotheses, observers would always set the ovoid to a circle on the screen and the data would all have an aspect ratio of 1. The symbols represent the results for one observer. Circles represent settings with monocular viewing through an aperture (frame not visible; MA), and triangles represent settings with monocular viewing without an aperture (frame visible; MF). Squares and inverted triangles represent settings with binocular viewing without apertures, squares for stimuli with rich pictorial information (BF+; an example of the stimulus is shown in Fig. 2c) and inverted triangles for stimuli with pictorial information removed (BF-).

We presented stimuli with rotated projections and observers viewed them with the display screen rotated by the same amount (Fig. 3a). Observers again judged aspect ratios of ovoids and slanted planes.

The predictions and results for the ovoid task are shown in Figure 3b. In the binocular-viewing condition with the frame visible, settings followed the surface predictions, particularly when the viewing angle was less than |45 deg|. In the monocular-viewing condition with an aperture, settings followed the retinal predictions. 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. With monocular viewing with an aperture, settings followed the retinal predictions; with binocular viewing, they followed the surface predictions. In conjunction with the results from the first experiment, these results show that adjustment for viewing obliqueness, when it occurs, is based on surface slant and not on pictorial information.

2.3 Experiment 3: Is the surface mechanism global or local?

Wide-angle pictures often produce distorted percepts even when the picture is viewed from the CoP (Pirenne, 1970; Kingslake, 1951; Zorin & Barr, 1995). The global surface-compensation hypothesis does not predict this effect: As long as the observer measures d_O, γ_{OP}, and S_{0M} correctly and makes a reasonable assumption concerning d_C, the percept will be undistorted. Importantly, the local surface hypothesis predicts that a viewer at the CoP will adjust the retinal image for points eccentric from the central surface normal, which is the wrong thing to do from a geometric standpoint. This would account for distorted percepts with wide-angle pictures viewed from the CoP.

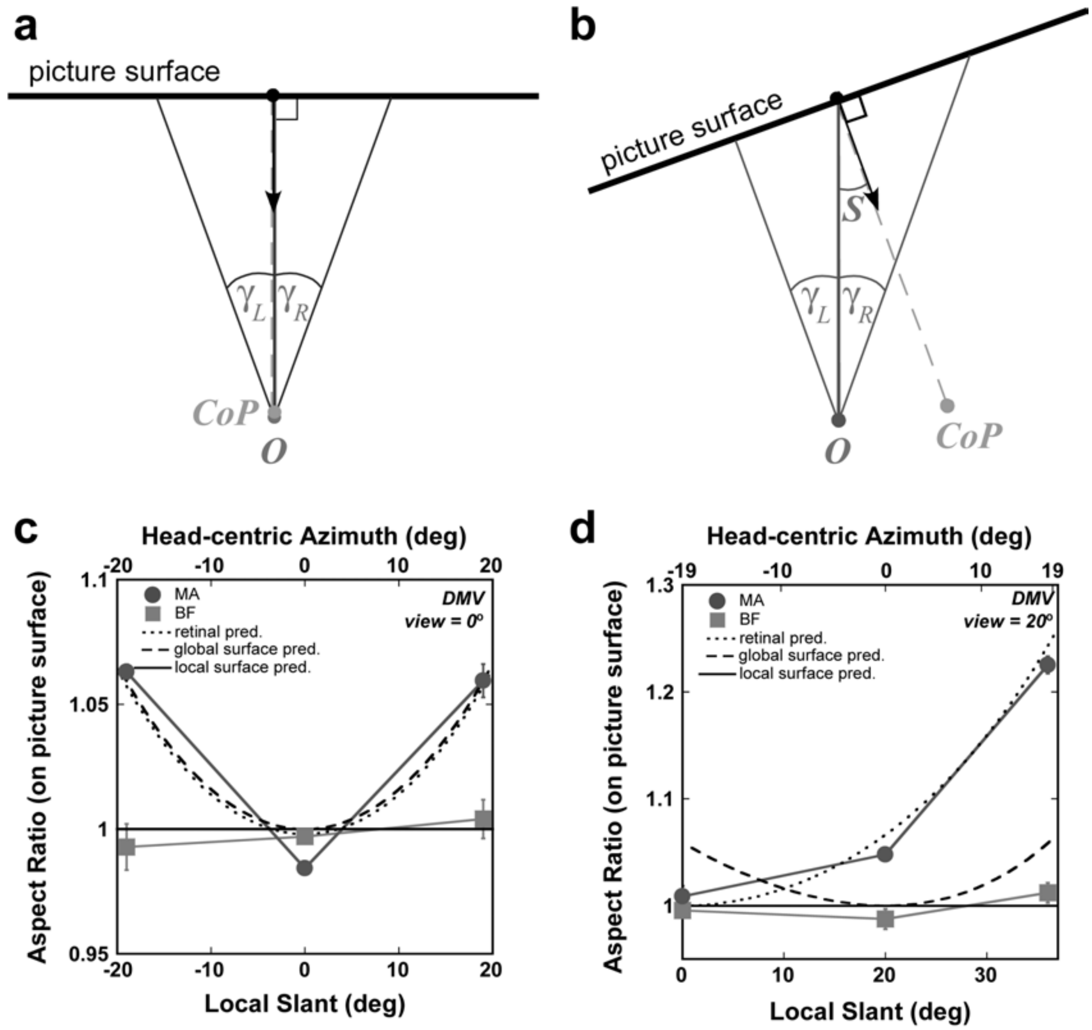

Figure 4. Stimuli, predictions, and results for the third experiment. a) and b) Plan views of the projection and viewing angles and the stimulus azimuths. The projection was always frontoparallel. The viewing angle with respect to the center of the display screen was either 0 (left panel) or 20 deg (right panel). Test ovoids were placed at head-centric azimuths (γ_L and γ_R) of –19, 0, or 19 deg when the viewing angle was 0 deg and at -20, 0, and 16 deg when the viewing angle was 20 deg. c) and d) Predictions and results for the ovoid task for viewing angles of 0 and 20 deg, respectively. Retinal predictions are represented by the dotted curves in both panels. They curve upward because the observer must set the eccentric ovoids to ellipses in order to generate a conical projection toward the eye. Predictions for global surface compensation are represented by the dashed curves in both panels; the global prediction is the same as the retinal prediction in the left panel, but differs in the right panel because the observation point and CoP were then different. Predictions for the local surface hypothesis are represented by the solid horizontal lines in both panels; they are horizontal at an aspect ratio of 1 because, according to this hypothesis, the observer will always set the ovoid to a circle on the screen. The symbols represent the results for one observer. Circles represent the settings with monocular viewing through an aperture (frame not visible; MA) and squares with binocular viewing and no apertures (frame invisible; BF).

The global and local hypotheses predict the same results for oblique viewing of the middle of the picture because for a point on the central surface normal, $\gamma_{OP} = 0$, so Eqn. 2 reduces to $S_{comp} = S_{0M}$, which is the same angle used by the local surface hypothesis. To tease them apart, we presented test objects at eccentric positions on the display screen. Observers judged the aspect ratios of ovoids presented at the center and 16-20 deg to the left and right of center while they viewed the display from the surface normal or 20 deg to the left. The conditions are schematized in Figure 4a,b, and the predictions and results are shown in Figure 4c,d. Settings followed the retinal prediction with monocular viewing with an aperture and the local surface prediction with binocular viewing and no apertures.
We conclude that the adjustments for viewing obliqueness are the consequence of the local surface mechanism. Such adjustments are triggered by the 3D orientation of the picture surface in the region of interest. As such, adjustments occur even when the viewer is at the CoP where the light field delivered to the eyes is geometrically correct. Such adjustments would cause the perceived distortions in wide-angle pictures (Pirenne, 1970).

3. DISCUSSION

3.1 Summary of current findings

With monocular viewing through an aperture, settings were based on the shape of the retinal image. With binocular viewing without apertures, they were based on the shape of the picture element on the surface; observers measured the slant of the picture surface in the region of interest and used the estimate to undo the local foreshortening. By using local surface slant, viewers adjusted the retinal image for viewing obliqueness in situations in which adjustment is unnecessary.

3.2 Quantitative model of adjusting for oblique viewing

We developed a model that can account for the observed behavior. The model has three parts. 1) It uses available evidence to determine the 3D orientation of the picture surface at each point of interest without contamination by 3D cues in the picture's contents. 2) The estimate of 3D orientation is then used to determine the foreshortening due to oblique viewing at each point of interest, and to adjust the retinal image to in effect undo foreshortening. 3) The 3D cues in the picture's contents are then interpreted.

1) Estimating the slant locally. We found that the local surface hypothesis provides the best account of how viewers take viewing obliqueness into account. We also argued that the picture viewer, in implementing adjustments based on local surface slant, must use signals that specify the slant of the picture surface and not the 3D structure of the depicted scene. Binocular disparity and the perspective of the picture frame are useful for this purpose because they are unaffected by the pictures' contents. How should the viewer determine the local surface slant? Bayes' Law prescribes how to weight evidence when making an estimate (Kersten, Mamassian, & Yuille, 2004). Its usage is consistent with the emerging account of how depth cues are used to form 3D percepts (Landy, Maloney, Johnston, & Young, 1995; Knill & Saunders, 2003). With no immediate consequences of the estimate (i.e., payoffs and penalties), the viewer should choose the maximum *a posteriori* estimate (MAP), which is the most probable given the image data and prior information. The MAP estimate for local slant derives from

$$p(S \mid i) \propto p(i \mid S)\, p(S) \qquad (4)$$

where S is surface slant and i is the input to the eyes. The first term on the right side is the likelihood function: the probability of observing various inputs given a particular slant presented to the eyes. The second term is the prior distribution: the probability of observing different slants at the eyes based on previous experience.

Assuming that disparity and the frame's perspective are conditionally independent (i.e., that their noises are statistically independent), we can write Eqn. 4 as

$$p(S \mid i_d, i_f) \propto p(i_d \mid S)\, p(i_f \mid S)\, p(S) \qquad (5)$$

where d and f refer to the disparity and frame cues, respectively. In our model, the visual system uses the maximum of the posterior distribution on the left side as its estimate of the local slant of the picture surface.

2) Undoing the foreshortening. In the model, the local slant estimate is input to Eqns. 3 such that the estimated foreshortening due to oblique viewing is undone. The veridicality of this step depends on the accuracy of the slant estimate from Eqn. 5.

3) Interpreting the picture. Once the estimated foreshortening has been undone, the viewer uses the remaining perspective information to interpret the picture's contents. The veridicality of the interpretation obviously depends on the accuracy of the preceding steps.

Some properties of the model. Whenever the standard deviation of either likelihood in Eqn. 5 is much smaller than that of the prior, the MAP estimate is close to the slant value presented to the eyes. Whenever the standard deviations of the likelihoods are much larger than the prior's, the MAP estimate is close to the peak of the prior distribution. The prior probability is proportional to cos(S) (defined from –90 to 90 deg) because steeply slanted surfaces project to small retinal images (Hillis, Watt, Landy, & Banks, 2004). The half cosine has a peak at $S = 0$ and standard deviation = ~40 deg. With binocular viewing, the standard deviation of the disparity likelihood at our distance of 45 cm is 6-10 deg, so the MAP estimate of slant when disparity is available will be very close to the slant presented to the eyes. This would yield complete or nearly complete invariance, as we observed. With monocular viewing through an aperture, the observer cannot estimate the surface slant reliably, so the standard deviation of the likelihood is very large and the MAP estimate approaches 0 deg, 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 deg| (Figs. 2c, 3b). At large slants the disparity gradient becomes large, and the ability to fuse the stimulus and estimate its slant is compromised (Burt & Julesz, 1980); in our model's framework, the standard deviation of the disparity likelihood increases, so the prior has greater influence, pushing the MAP slant estimate toward 0 deg.

According to our model, only binocular disparity and the perspective of the picture frame should be used to estimate the local slant of the picture surface. We observed a large effect of disparity, but only a small effect of frame visibility (Figs. 2c, 3b). The small frame effect could have been the consequence of stimulus and task characteristics and/or the visual system's assumptions about the frame's objective shape. 1) Regarding the stimulus and task, observers generally fixated and attended to the target ovoid in the center of the display. They may not have picked up the slant information from the frame because it fell in the retinal periphery. 2) Regarding the frame's objective shape, using the frame to estimate the slant of the picture surface depends on the assumption that its objective shape is known (i.e., rectangular). Uncertainty about its objective shape may have reduced the frame's effect. In the model, the failure to attend to the frame or uncertainty about its objective shape would be expressed as a large standard deviation for the frame likelihood in Eqn. 5, which in turn would yield little effect on the MAP slant estimate.

3.3 Do adjustments for oblique viewing require special pictorial mechanisms?

Investigators have debated whether picture viewing requires special mechanisms (Hagen, 1970; Rosinski & Farber, 1980). Our data and analysis suggest a new way to evaluate this issue. As we said earlier, the picture viewer at an arbitrary viewing position must treat separately two perspective effects at the retina: perspective due to viewing obliqueness and perspective due to the picture's contents. We believe that both parts are manifestations of everyday visual functions and not of special mechanisms for pictures. Adjusting for obliqueness is manifest when a person reaches to pick up an object. Consider a book lying ahead on the desk. It is slanted relative to the line of sight, so the person must first estimate the book's width from the foreshortened retinal image in order to open the hand by the right amount. People are very good at this (Mamassian, 1997); they exhibit shape constancy (Epstein & Park, 1963). Interpreting the picture's contents after obliqueness adjustment is a manifestation of inferring 3D layout from the variety of depth cues present in pictures.

The part that is special to pictures is the need to segregate perspective effects according to their cause, so that the interpretation of the 3D layout of the picture's contents is not contaminated by the perspective distortions caused by viewing obliqueness.

4. CONCLUSION

We investigated the perception of pictures viewed from the wrong place. We found that viewers use a local slant mechanism to estimate foreshortening due to viewing obliqueness and then adjust the retinal image to undo the foreshortening. This mechanism yields reasonable invariance except for wide-angle pictures where it produces perceived distortions. We presented a quantitative model that states how the various measurements and adjustments are made. The model's behavior is quite consistent with the percepts of picture viewers.

With respect to where to sit at the movies, our findings suggest that a binocular viewer, sitting sufficiently close to the screen to estimate its slant from disparity, will not experience noticeable distortions because of the ability to adjust for obliqueness of view. However, a variety of circumstances should make distortions due to viewing obliqueness noticeable: sitting at the edge of the theater such that slant of the movie screen exceeds 45 deg, or sitting obliquely and far from the screen such that slant from disparity is unreliable. Furthermore, viewers with deficient binocular vision are more likely to experience distortions from oblique viewpoints.

5. REFERENCES

Burt, P. & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science* **208**, 615-617.

Caprile, B. & Torre, V. (1990). Using vanishing points for camera calibration. *International Journal of Computer Vision* **4**, 127-140.

Cutting, J.E. (1987). Rigidity in cinema seen from the front row, side aisle. *Journal of Experimental Psychology: Human Perception & Performance* **13**, 323-334.

Epstein, W.P. & Park, J.N. (1963). Shape constancy: Functional relationships and theoretical formulations. *Psychological Bulletin* **60**, 265-288.

Gårding, J. (1992). Shape from texture for smooth curved surfaces in perspective projection. Journal of Mathematical Imaging and Vision **2**, 329-352.

Gershun, A. (1939). The light field. *Journal of Mathematics and Physics* **23**, 51-151.

Gibson, J.J. (1950). *The Perception of the Visual World*. Houghton-Mifflin, Boston, USA.

Goldstein, E.B. (1987). Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *Journal of Experimental Psychology: Human Perception & Performance* **13**, 256-266.

Gombrich, E.H. (1960). *Art and Illusion*. (Princeton University Press, Princeton, USA).

Hagen, M.A. (1976). Influence of picture surface and station point on the ability to compensate for oblique view in pictorial perception. *Developmental Psychology* **12**, 57-63.

Hillis, J.M., Watt, S.J., Landy, M.S., & Banks, M.S. (2004). Slant from texture and disparity cues: optimal cue combination. *Journal of Vision* **4**, 967-992.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian Inference. *Annual Review of Psychology* **55**, 271-304.

Kingslake, R. (1951). *Lenses in Photography: The Practical Guide to Optics for Photograhpers*. (Case-Hoyt).

Koenderink, J.J. & van Doorn, A.J. (2003). Pictorial space. *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space*. Hecht, H., Schwartz, R., & Atherton, M., Eds. (MIT Press, Cambridge, USA).

Knill, D.C. & Saunders, J.A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research* **43**, 2539-2558.

Kubovy, M. (1986). *The Psychology of Perspective and Renaissance Art*. (Cambridge University Press, New York).

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.

Leonardo da Vinci. (1970). *The Literary Works of Leonardo da Vinci*. J.P. Richter, Ed. (Phaidon, London, England).

Mamassian, P. (1997).Prehension of objects oriented in three-dimensional space. *Experimental Brain Research* **114**, 235-245.

Meister, R. (1966). The iso-deformation of images and the criterion for delimitation of the usable areas in cine-auditoriums. *Journal of the Society of Motion Picture & Television Engineers* **75**, 179-182.

Perkins, D.N. Compensating for distortion in viewing pictures obliquely. *Perception & Psychophysics* **14**, 13-18 (1973).

Pirenne, M.H. (1970). *Optics, Painting and Photography*. (Cambridge University Press, Cambridge, England).

Rosinski, R.R. & Farber, J. (1980). Compensation for viewing point in the perception of pictured space. *The Perception of Pictures*. M.A. Hagen, Ed. (Academic Press, New York, USA).

Rosinski, R.R., Mulholland, T., Degelman, D., & Farber, J. (1980). Picture perception: An analysis of visual compensation. *Perception & Psychophysics* **28**, 521-526.

Sedgwick, H.A. (1991). The effects of viewpoint on the virtual space of pictures. *Pictorial Communication in Virtual and Real Environments*. Ellis, S.R., Kaiser, M.K., & Grunwald, A.C., Eds. (Taylor & Francis, London, England).

Wallach, H. & Marshall, F.J. (1986).Shape constancy in pictorial representation. *Perception & Psychophysics* **39**, 233-235.

Zorin, D. & Barr, A.H. (1995). Correction of geometric perceptual distortions in pictures. *Proceedings of SIGGRAPH 1995, ACM SIGGRAPH* **14**, 257-264.