



OPEN Viewers perceive shape in pictures according to per-fixation perspective

Daniel Martin^{1,2}✉, Diego Gutierrez¹, Belen Masia¹, Stephen DiVerdi² & Aaron Hertzmann²

How viewers interpret different pictorial projections has been a longstanding question affecting many disciplines, including psychology, art, computer science, and vision science. The most-prominent theories assume that viewers interpret pictures according to a single linear perspective projection. Yet, no existing theory accurately describes viewers' perceptions across the wide variety of projections used throughout art history. Recently, Hertzmann hypothesized that pictorial 3D shape perception is interpreted according to a separate linear perspective for each eye fixation in a picture. We performed four experiments based on this hypothesis. The first two experiments found that viewers consider object depictions as more accurate when an object is projected according to its own local linear projection, rather than one consistent with the rest of the picture. In the third experiment, viewers exhibited change blindness to projections in peripheral vision, suggesting that perception of shape primarily occurs around fixations. The fourth experiment found surface slant compensation to be dependent on fixation. We conclude that pictorial shape perception operates according to per-fixation perspective.

When viewing a photograph or realistic painting, a viewer may interpret 3D shape and space in the implied scene, whether real or imaginary. In this work, we use “shape” to refer to the geometric properties of an object, i.e., the spatial arrangement of its boundaries or surface, which remain unchanged by rigid transformations. The geometric relationship between a 3D shape and a 2D depiction can be described in terms of a *projection*. Can we describe viewers' 2D interpretations of 3D depictions in terms of projections?

This question is complicated by the fact that artists throughout history have employed many different projection techniques^{1–3}. Some of these, like linear perspective and parallel projections, can be created with explicit drawing systems. Current cameras, visualizations and video games offer other parametric projections for various uses, like stereographic and equirectangular projection (Fig. 1). Moreover, from cave painting to contemporary art, many styles of depiction do not seem to follow any strict system—instead, artists seem to adapt 2D arrangements and layout to scene content and artistic goals⁴. Explaining how such depictions convey 3D shape and space seems to defy any simple projection formula. Computational photography techniques, including content-aware and multiperspective projections^{4–13} offer techniques both for reducing perceived distortions and for replicating projections in art, without following any simple, unique projection formulae.

Despite this diverse history, studies of projection overwhelmingly focus on *linear perspective*. Geometrically, linear perspective projection can be defined by a COP and a picture plane (Fig. 2a). Each 3D scene point projects to the picture along a ray through the COP. In art history, advocates for linear perspective drawing techniques presented them as a scientific, rational approach to projection^{14,15}, because, in theory, a viewer located at a picture's COP sees the same light as though viewing the original scene. Yet, throughout art history, we find that many artists mastered linear perspective techniques but applied them selectively, or rejected them entirely^{1–3,16,17}. Even famous practitioners of realistic linear perspective do not strictly obey its rules in wide-angle depictions, exhibiting, for example, systematic deviations for faces and spheres^{16,18,19}, and for objects distant from the viewer^{20,21}. A reason for this may be that wide-angle linear perspective pictures are not normally viewed from the COP^{17,22}, thereby producing well-known perceptual distortions such as marginal distortion and perspective distortion^{22,23}. Even when viewed from the COP, marginal distortion still occurs in binocular viewing^{19,22}. Moreover, the hypothesis that viewers interpret linear perspective pictures as though viewing from the COP²⁴ has been rejected by extensive experimental studies^{23,25–29}.

How might we explain these phenomena—not just in linear perspective, but in broader classes of projections? How do viewers perceive shape, given a certain projection? One key observation is that, under normal viewing conditions, a viewer makes multiple fixations and eye movements to understand a picture. Due to the limits of peripheral vision^{31,32}, a viewer obtains detailed pictorial information for only a few degrees of visual angle

¹Universidad de Zaragoza, I3A, Zaragoza, Spain. ²Adobe Research, California, USA. ✉email: danims@unizar.es



Fig. 1. Four common projection systems, used for various visualization and artistic purposes. Each one conveys different shape percepts and perceived distortions. Top left: A linear perspective photograph taken with a smartphone in ultrawide zoom, showing marginal distortions in the sphere and the face. Top center: An isometric projection, common in design and videogames, allowing the depiction of large-scale scenes without apparent distortions. Duplicates of the same object all have the same appearance and size, regardless of their locations. (Illustration by Giulia Calistro.) Top right: A stereographic photograph, which distorts the surface of the Earth to appear like a “little planet”. (Photograph by Rich Radke.) Bottom: A pushbroom panorama, which combines multiple sequential cartographic images within the same axis, with one center-of-projection for each column of pixels. (Panorama by Román and Lensch⁸).

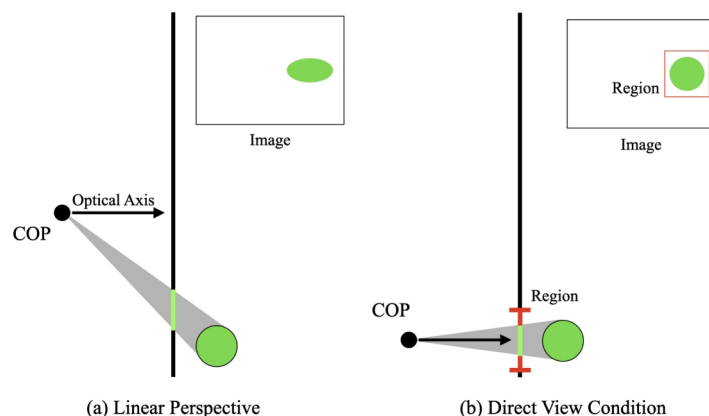


Fig. 2. Possible projection models for viewers' implicit assumptions in shape perception. **(a)** Conventional linear perspective uses a single COP for the entire picture. The original vantage-point compensation hypothesis²⁴ states that viewers interpret pictures according to an estimate of the picture's true COP, but extensive experimental evidence—combined with the fact that many pictures do not have COPs—contradict this hypothesis^{22,23}. As depicted in the inset, linear perspective causes objects in the periphery to appear distorted, a phenomenon known as *marginal distortion*. **(b)** In the Direct View Condition^{19,30} hypothesis, image regions appear undistorted when they are consistent with a local-linear perspective that has the COP in front of the object. The inset shows a picture for which the region containing the green object is consistent with a local-linear perspective with effective COP in front of the object. Therefore, in this case, it does not appear distorted..

around each fixation, while much coarser information is obtained from peripheral vision. Another observation is that the shape of an object in a picture rarely depends on the context around it: cropping an object from a picture, placing it into a different location in a collage, or even placing it into a picture with a different projection (Fig. 3) does not affect the perceived shape of the object, apart from highly-ambiguous shapes. Based on these observations and others, Hertzmann¹⁹ recently proposed a series of hypotheses about the projections that describe viewers' interpretations of 3D shape and structure in pictures. Specifically, Hertzmann hypothesized that viewers infer shape primarily in the area around each fixation, in a manner consistent with a locally-linear perspective centered at the fixation, and shapes inconsistent with such projections will appear distorted.

Inspired by these hypotheses, here we set out to empirically test whether foveal vision and local-linear perspectives could explain pictorial shape perception. We performed four experiments. Experiment 1 evaluates,

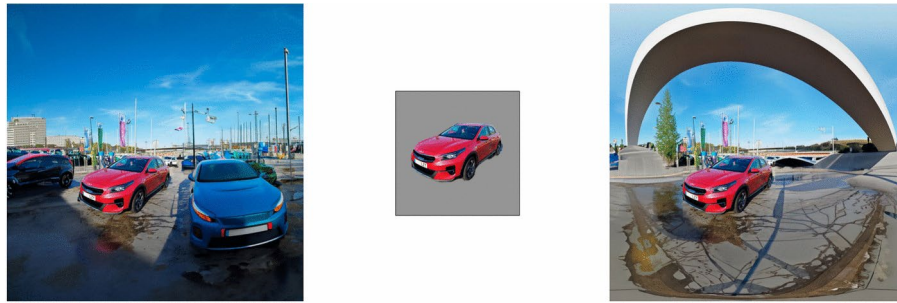


Fig. 3. The hypotheses in this paper explore shape perception as a local phenomenon, rather than one that depends on joint inference of projection parameters, such as projection model or camera focal length. This figure shows three different images of the same car within very different contexts, suggesting different projections. The car appears the same, illustrating that perception of shape is local, not dependent on the rest of the picture.

under a range of common projections, whether viewers prefer globally-consistent projections, or local-linear projections instead. The latter would be consistent with what the *Direct View Condition* (DVC) hypothesis^{19,30} posits: that depicting an object according to a linear perspective centered on the object guarantees that the object will not appear distorted, regardless of the contents of the rest of the picture. This contrasts with conventional linear perspective, which requires a single COP for an entire picture. Experiment 2 evaluates absolute distortion judgments instead of comparative, and expands the range of configurations tested for global and local linear perspectives. In Experiment 3, we test whether viewers are sensitive to shape details in peripheral vision, by testing for change blindness. Experiment 4 tests whether surface slant compensation at fixations can explain the interpretation based on a locally-linear perspective.

Experiment 1: comparing perceived distortion among projections

Our first experiment tests which projections, when applied to shapes, cause them to appear distorted. We consider two classes of projections: projections applied consistently to an entire picture, including conventional linear perspective, and locally-linear projection.

Theories based on linear perspective have been the most prominent. For example, in the *vantage-point compensation* hypothesis, viewers interpret pictures according to linear perspective projection, while accounting for the mismatch between the viewer's position and COP in some way²⁴. If this hypothesis is correct, then pictures created with any linear perspective will not appear distorted. However, previous evidence indicates that viewers do not compensate for viewing position^{22,23,25–29}, and linear perspective creates marginal distortion. A few authors have proposed various curvilinear projection systems as being more “perceptually accurate” or “natural” than linear perspective, or as better describing projections in art history^{17,20,33–35}. These alternative projections each have some advantages, but none has received widespread acceptance. Since different projection schemes have been used in different situations, and no single scheme avoids distortion in all contexts, one might predict that viewers expect *global consistency*, i.e., shapes should be consistent with *some* global projection technique.

In contrast, the *Direct View Condition*^{19,30} (DVC) predicts that perceived distortion does not depend on the rest of the picture. Instead, to appear undistorted, an object should appear as it would at the center of a normal-FOV linear-perspective picture. Otherwise, it will appear distorted (Fig. 2b). For example, at the center of a linear perspective picture, a sphere projects to a circle. Consequently, the DVC predicts that spheres will appear undistorted if and only if they are depicted with circular outlines—regardless of how the picture was created, or where the sphere appears in it. Spheres often project to approximately circular shapes in many parts of linear perspective pictures, and so they do not appear distorted. They only appear distorted in the margins, and the more oblong their outlines, the more distorted they appear. Spheres also project to circles in other projections, including stereographic and parallel projections, and so spheres never appear distorted in these projections. A painting that has no explicit projection formula can make spheres look undistorted if they are depicted as circular. The same reasoning applies to any other kind of shape or region, e.g., the DVC predicts that straight lines must project to straight lines to appear undistorted; symmetric shapes should project to symmetric shapes; and so on. Here, “normal-FOV” refers to typical fields-of-view that avoid extreme depth compression and expansion, e.g., as with 50 mm focal lengths, which correspond to approximately 39° horizontal FOV²³. We have phrased the DVC slightly differently from previous authors in order to emphasize that the DVC describes the contents of a picture region, and does not depend on the viewer's perceptual process.

The DVC predicts commonplace depiction techniques, e.g., straight lines should project to straight lines and spheres should project to circles^{10,30}; it predicts preferred depictions for any other object as well. In this hypothesis, shape perception does not depend on picture contents outside of a small object region, and it does not describe the relationships of objects in separate regions. In the DVC, objects that appear undistorted in a photograph would appear equally undistorted if depicted the same but arranged in a more peculiar manner, as in, say, a Giorgio de Chirico painting or the Ponzo illusion.

In this experiment, we consider a specific prediction of the DVC: if a shape is projected with a *local-linear* projection, then viewers perceive it as undistorted. We compare this prediction with *global consistency*: if the whole picture is projected with a single consistent projection, then viewers will judge the shapes as undistorted.

Results

To compare the two predictions, we constructed a virtual 3D scene, and rendered it with five different parametric projections. We chose them to cover a variety of projections that viewers may have already been exposed to, including wide-angle perspective (e.g., smartphone cameras), isometric projection (e.g., videogames) and stereographic projection (e.g., point-of-view (POV) recordings). For each projection, we render a familiar object prominently located in the picture (Fig. 4). For each picture, we created two conditions. In one condition, the familiar object was projected in the same manner as in the rest of the picture (*globally-consistent*). In the second condition, the familiar object was projected with its own *local-linear perspective*: a linear perspective projection, centered on the object⁴. We repeated this with two different familiar objects (a volleyball and a bus). For every pair of projection and object, images for the two conditions (*globally-consistent* and *local-linear*) were presented simultaneously side-by-side, and the following question was posed: “Which object looks more like a volleyball?” or “Which object looks more like a bus?” as appropriate. These conditions were chosen as cases that could discriminate between local projection versus theories based on global projections. We additionally used a sentinel case (Fig. 4, bottom row) to check that participants performed the task correctly.

We collected twenty-four responses from each of the twenty-five participants (two repetitions per pair), leading to a total of 600 judgments. We first studied the difference in the number of votes between the globally-

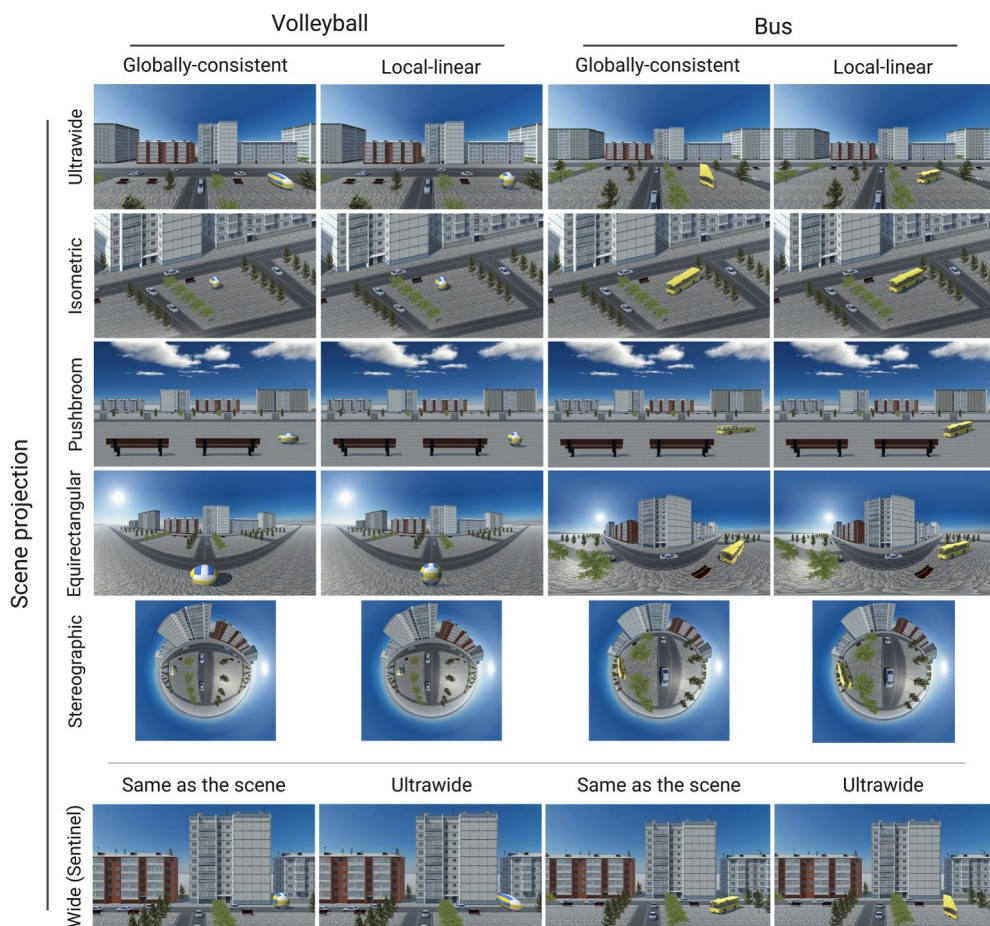


Fig. 4. Pairs of scenes used in Experiment 1. We have rendered the same scene using five different projections (plus a sixth used as sentinel, last row), including both linear and curved perspectives that cover a variety of projections used in everyday life, such as wide-angle perspectives in smartphone cameras or isometric perspectives for videogames, among others. Then, for each projection, we have rendered one image depicting a familiar object—either a volleyball (left columns) or a yellow bus (right columns) both with the same projection as the scene (i.e., globally consistent) and with local-linear perspective. We asked participants to judge “Which object looks more like a volleyball/bus?”, respectively. This tests whether viewers perceive shape according to a global consistency theory, or whether a different, local-linear theory, such as the Direct View Condition³⁰ hypothesis, which predicts a preference for the familiar objects rendered following local-linear perspective, better fit their perception. The bottom row shows sentinel tasks, in which the local-linear projection is replaced with an ultrawide projection, which is both inconsistent and distorted.

consistent and local-linear conditions. Participants' responses show a clear preference for the familiar object rendered with locally-linear projection over the globally-consistent one in four of the five projections tested (Fig. 5a). Since our data are not normally distributed (Shapiro-Wilk test, $p < 0.05$), we conducted a Wilcoxon signed-rank test across projections to test significance. We found that differences in preference in the wide angle, pushbroom, and equirectangular projections were significant ($p < 0.05$). These results contradict the global consistency prediction, and match the local-linear prediction, and thus support the DVC. For the sentinel case, 99 out of the 100 responses were in favor of the globally-consistent one, indicating that participants were focusing correctly on the task. For the isometric projection, no significant difference between conditions was found ($p > 0.05$). This result is still consistent with the DVC, since the object appearance is very similar between the two conditions.

For the volleyball, the projected aspect ratio (the ratio of width to height) provides a simple measure of distortion, since, according to the DVC, a sphere should project to a circle. Greater deviation from the uniform aspect ratio should correspond to greater perceived distortion. The relationship between projected aspect ratio, and votes for the projection, are plotted in Fig. 5b). These correlations (Pearson's $r = 0.95$, Spearman's $\rho = 0.89$) show that participants tended to have a stronger preference for the local-linear condition when the globally-consistent one deviated most from a uniform aspect ratio. This supports the idea that greater deviations from local-linear perspective lead to a greater perception of distortion.

Experiment 2: perceived level of distortion

Our second experiment extended the first, using a similar experimental setup. In this experiment, we obtained viewers' absolute judgments of distortion in individual pictures, in addition to comparative judgments. Additionally, we tested a range of configurations in order to see how perceived distortion varies with field-of-view. Here we experimented solely with local-linear projection and global-linear perspective. We particularly explore a continuum between *wide-angle* and *ultrawide* FOVs, exploring whether preference for a local-linear projection over a globally-consistent one holds across different degrees of distortion.

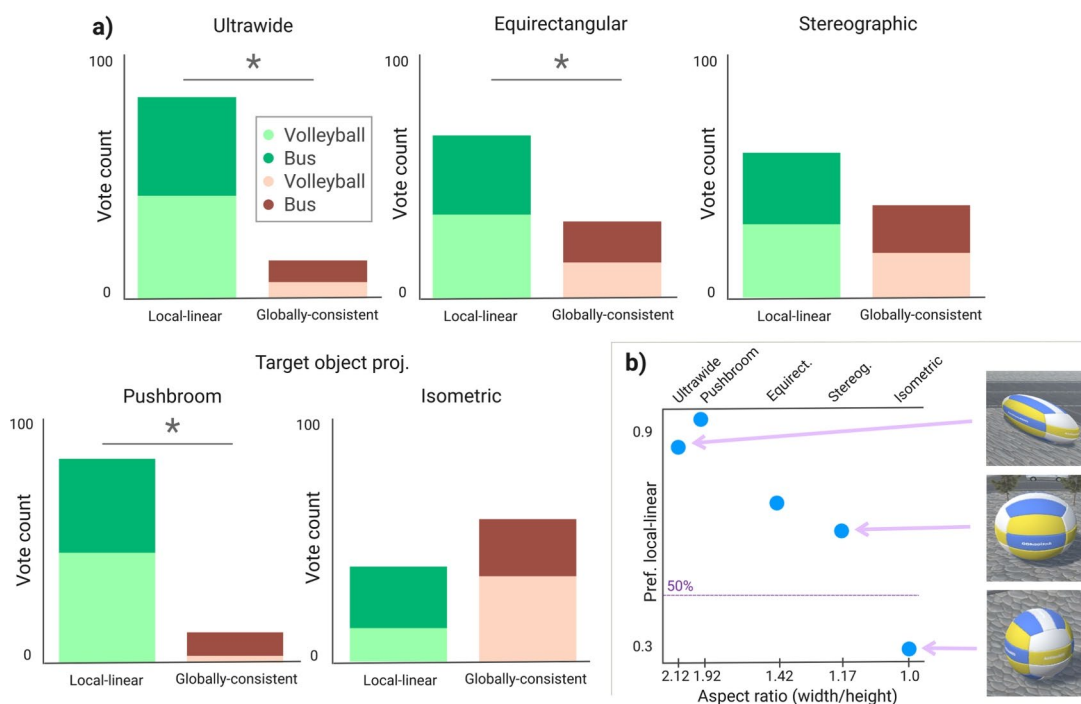


Fig. 5. Results from Experiment 1. **(a)** Participants' preferences for each of the five projections studied (one per plot). Vote counts with a statistically significant difference (Wilcoxon signed-rank test with p -value < 0.05) are marked with an asterisk. In four of the projections, participants showed a preference for the pictures with the familiar object rendered following local linear perspective (three of them statistically significant), which seems to contradict the global consistency theory, and indeed supports the predictions from the Direct View Condition (DVC)³⁰. **(b)** Participants' preference for the local linear condition w.r.t. the image-space aspect ratio of the volleyball rendered with the same projection as the scene (globally-consistent case). The more distorted the sphere (i.e., the more different the projection was w.r.t. linear perspective), the less probable it was for participants to choose it (Pearson's $r = 0.95$, Spearman's $\rho = 0.89$); this seems to suggest that people tend to prefer those projections where the object is closest to its normal projection, again contradicting global consistency and supporting DVC.



Fig. 6. Stimuli used in Experiment 2. Left: We render a scene with nine different fields of views (FoV), ranging from *wide-angle linear perspective* (60°) to *ultrawide linear perspective* (140°) in steps of 10° . Right: Similar to Experiment 1, for each of those images, we render the same scene but with a prominent object (i.e., a yellow bus) following a local-linear perspective. We asked participants to judge “Which object looks more like a bus?” We set the bus in two different orientations (rows). We then ask people to provide evaluate the distortion using a 1-5 Likert scale. Similar to the previous experiment, this tests whether people’s judgments of perceived distortion is consistent with the global-consistency prediction, or the local-linear prediction of the DVC..

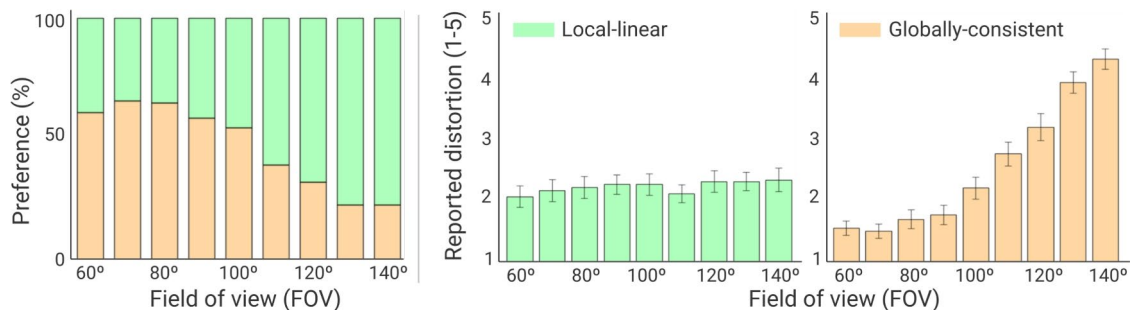


Fig. 7. Results from Experiment 2. Left: Participants’ preferences for each of the nine fields-of-view evaluated (see Figure 6). We found a positive correlation between preference for *local-linear* projections (green) and the increase in FOV (Pearson’s $r = 0.94$, Spearman’s $\rho = 0.95$), suggesting that as FoV increases (i.e., distortion becomes more evident), viewers’ tend too better prefer the local-linear condition. Right: Participants’ rating (1-5 Likert scale) of the distortion of the yellow bus for all FOV in both *local-linear* (green) and *globally-consistent* (brown) projections. The error bars show the standard error of the mean (SEM). Our results show that people clearly report higher distortion in the images conforming to a single global system, but do not in the local-linear ones. Both results contradict the global-consistency hypothesis and support local-linear theories such as the Direct View Condition (DVC).

Results

In this experiment, we evaluate whether the preference for local-linear projections over globally-consistent ones occurs in projections in which distortion is less extreme. We resort to the same 3D scene depicting a yellow bus as in the previous experiment, and render it using nine different FOVs, ranging from *wide-angle linear perspective* (60°) to *ultrawide linear perspective* (140°) in steps of 10° . For each picture, we created a *globally-consistent* and a *local-linear perspective* conditions (Experiment 1) and presented them simultaneously side-by-side, asking participants “Which object looks more like a bus?”. We did this for two different bus orientations (see Fig. 6).

We collected fifty-four responses from each of the twenty-five participants (three repetitions per pair), leading to a total of 1,120 judgments. We found a positive correlation between preference for local-linear projections and the increase in FOV (Pearson’s $r = 0.94$, Spearman’s $\rho = 0.95$, see Fig. 7a), indicating that viewers judge local-linear perspectives as less-distorted than global consistency, confirming that the results found in Experiment 1 holds for cases with less extreme distortions, and contradicting the global-consistency predictions.

After completing the fifty-four comparison trials, we asked participants to evaluate the level of distortion of each bus separately on a 1-5 Likert scale: “In a scale of 1 (not distorted) and 5 (very distorted), how distorted do you think the yellow bus is?”. Results (Fig. 7b) show that for images following the *local-linear* condition, distortion judgments are consistently low regardless of the field-of-view of the scene, while they significantly increase for the case of *globally-consistent* conditions. This suggests that the perception of shape and distortion only depends on the depiction of the bus itself, and is not affected by any distortion on the scene around, which contradicts global-consistency theories and supports theories such as the DVC, which claims that “the greater

the deviation in appearance from the appearance at the center of a linear perspective picture, the greater the appearance of distortion.”¹⁹

Experiment 3: projection in peripheral vision

In the first two experiments, viewers showed a preference for pictures with local linear perspective, as compared to projections that are globally-consistent across the picture. What might explain this finding?

The properties of foveal and peripheral vision could provide some explanation. Peripheral encodings represent visual information using summary statistics that discard some spatial information^{31,36}. The information loss in peripheral encodings can cause viewers to be oblivious to entire objects, particularly in the presence of crowding. This fact provides an explanation for many change blindness experiments³²: viewers are sometimes unaware of objects in peripheral vision. Hence, in order to test viewers’ awareness of projections in peripheral vision, we performed a change blindness experiment.

Results

In this experiment, we predicted a change blindness effect: viewers will be oblivious to some changes to projection in peripheral vision. Specifically, participants were shown a rendering of a 3D scene with a large object in each corner of the picture (Fig. 8(top left)) in a display within a virtual reality (VR) environment. We used VR to simulate the experience of showing a picture on a wall while having convenient eye-tracking. Participants were instructed to fixate on the center of the picture and then told to move their eyes to a specific corner of it. When a saccade was detected by an eye tracker, the system simulated a blink briefly, by blanking the display with a neutral gray^{37–39}. During the blink, the projection of an object in the relevant picture quadrant changed (Fig. 8(center)). As controls, we also included trials with other kinds of modification, such as changing object color (Fig. 8(right)), as well as trials where no modification occurred. Viewers were instructed to indicate whether they had detected any change in the image. If viewers perceive the object’s appearance during the initial (central) fixation and remember the appearance, then they should detect these changes. On the other hand, if they do not perceive details of shape in peripheral vision, then they should not detect a change.

Figure 9 shows the detection ratio for each image modification, i.e., the percentage of times that each modification was detected. On average, viewers detected projection modifications only 16.7% of the time. Modifications in the bottom-left and top-right corners were never detected, and only two participants detected the other two. Participants were slightly better at detecting non-projection modifications (detection ratio of 25%), but still well below chance level. Only a color modification (depicted in Fig. 8(bottom left)) was detected half of the time.

One possible explanation for these results is that, while fixating on the picture center, viewers were unaware of the precise shapes of the objects in the corners. This seems consistent with the known limitations of peripheral vision³¹. A second possible explanation is that viewers did not remember the image’s appearance between fixations well enough to compare. Both interpretations support the idea that, in normal viewing, shape perception depends primarily on foveal vision.

In a post-trial survey, we asked participants whether they had seen certain pictures during the experiment. The survey included both actual stimulus pictures that had been shown to them in the experiment, and a *composite* picture (Fig. 8(bottom left)) that they had never actually been shown. The composite comprised four quadrants taken from pictures that the viewers had seen (but never together). All participants reported having seen the composite, despite the fact that this picture had never been shown (see Fig. 9(right)). In contrast, all but one participant (83.3%) reported *not* having seen the *initial* image, even though it was shown to all participants at the start of every trial. This can be explained by the fact that no participant had fixated on any corner of this image. In a sense, the composite picture seems to have summarized viewers’ visual experience, since it combines

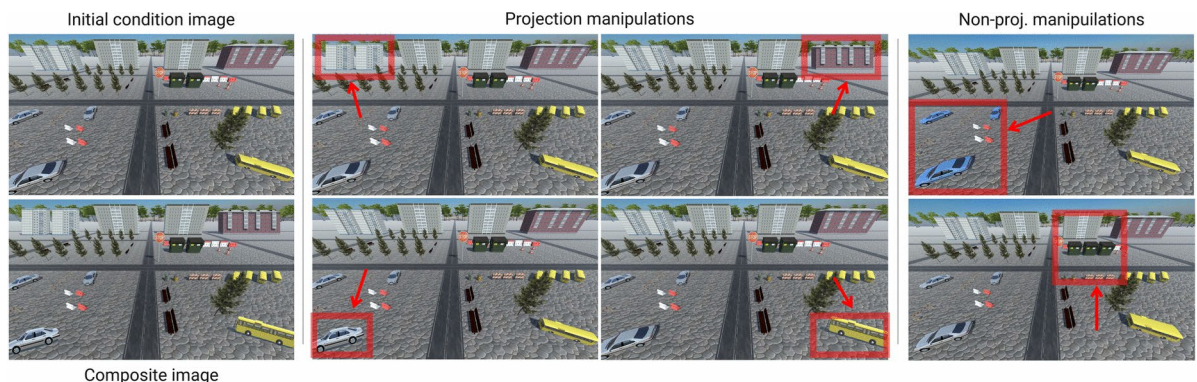


Fig. 8. Stimuli used in Experiment 3. Top left: *Initial condition* scene. Each test condition involves a modification to an object in one of the four corners of the picture. In four cases, the shape is projected with a local-linear projection, e.g., as shown on the center. In four other cases, non-geometric modifications are performed, e.g., right. Besides, in a post-experiment survey we show participants a *composite* image (bottom left) containing all four projection manipulations.

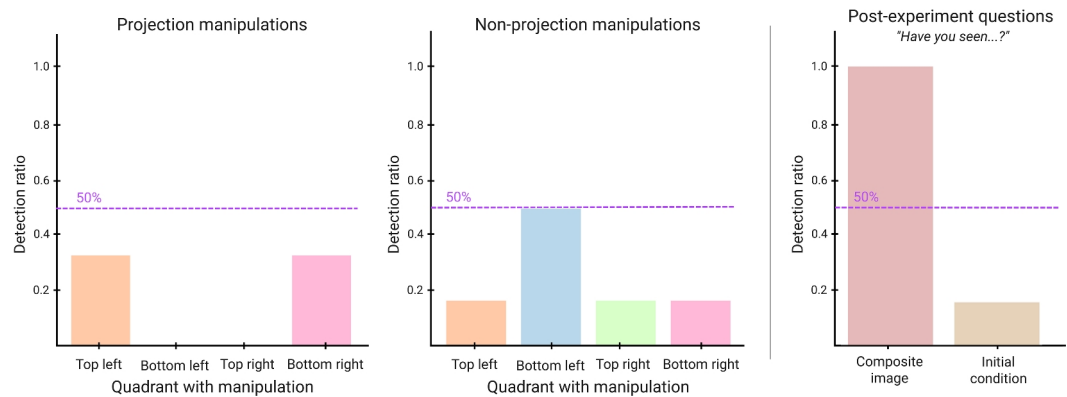


Fig. 9. Results from Experiment 3. Left: Detection ratio for each of the four modifications related to changes in projection. Note that each change remained mostly undetected, suggesting that people were not able to detect that something had changed in the periphery. Center: Detection ratio for each of the non-projection manipulations. We designed those manipulations to be closer to the center and slightly more overt, although participants still exhibited a low detection ratio. Right: Results for our post-experiment questions. We asked participants if they had seen two pictures, one picture containing the four projection modifications at the same time (*composite*), and the *initial condition* image. All participants reported having seen the composite picture, which they had not been shown, whereas most reported not having seen the initial picture, which they had been shown at the start of every trial.

all the picture regions that viewers ever fixated upon, whereas each picture they were actually shown included regions that viewers never fixated on.

These results indicate that viewers perceived and remembered the shapes seen in foveal vision, while being unaware of (or unable to remember) shape details in peripheral vision, at the granularity that we tested. Visual experience of a picture may be best thought of as a composite of the experiences of individual fixations, rather than as taking in a single picture all at once. However, this composite may not be globally-consistent.

Experiment 4: slant compensation

In the fourth experiment, we sought to understand the role of slant compensation^{22,24}: viewers' mental adjustment for the slant angle of a picture surface. In the absence of slant compensation, picture contents would appear to continually skew with head movement—a sphere in a picture would look ellipsoidal from most viewpoints. Hence, objects would appear distorted from most viewpoints. Yet, pictured objects typically look the same across a range of viewpoints. If viewers perceive shape according to a local-linear perspective (as predicted by the DVC), then each pictured object should be perceived as though viewed frontally, regardless of the actual picture surface slant. This would require compensating for slant at the depicted object's position within the picture.

Indeed, Vishwanath et al.²² found evidence for *Local Slant Compensation*: a viewer compensates for foreshortening due to picture slant at a point of interest. This compensation functions as though mentally stretching a depicted object's appearance, in proportion to the degree of foreshortening. Since Experiment 3 showed limited shape perception outside of foveal vision, we next studied whether peripheral vision affects slant compensation.

We hypothesized that Local Slant Compensation predicts pictorial shape perception in foveal vision, but not in peripheral vision. If true, this would provide additional evidence for the dependence of shape perception on fixation. If, instead, Local Slant Compensation predicts pictorial shape perception well in both foveal and peripheral vision, then the importance of fixations in perceiving shape would seem more limited.

We performed an experiment based on Vishwanath et al.'s compensation experiments²², with the addition of controlled eye fixations in a binocular virtual reality (VR) environment. In each trial, a participant faced the center of a picture display (Fig. 10a), showing a picture containing a red ellipsoid. The red ellipsoid appeared either to the left or the right of the picture center so that the ellipsoid was viewed at a slant $\theta = \pm 20^\circ$, either to the left or to the right. Participants were instructed to fixate with a specific eccentricity ξ . We set ξ to six different values for each θ , so that to have an angle between both $\alpha = \theta - \xi \in \{0^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ, 50^\circ\}$ (Fig. 11a). When $\xi = \theta$, the participant fixated on the ellipsoid, otherwise, the ellipsoid appeared in peripheral vision. In each condition, the ellipsoid appeared on a slanted picture surface, and so foreshortening narrowed the ellipsoid in the retinal image. For example, a sphere on a foreshortened picture would project to a narrow ellipse in the retinal image, instead of a circle. Slant compensation may correct for this foreshortening²². We constructed the stimuli so that different depicted objects were visible in peripheral vision, using bright red against backgrounds with darker colors. The participants did not report any difficulty in performing the task.

Results

In each condition, the participant must judge whether the ellipsoid is too wide or too narrow to be a perfect sphere. Differences in perceived aspect ratios of shapes for different slants (if any) provide evidence for the type of compensation performed. We gathered 168 responses per participant in total. For each combination

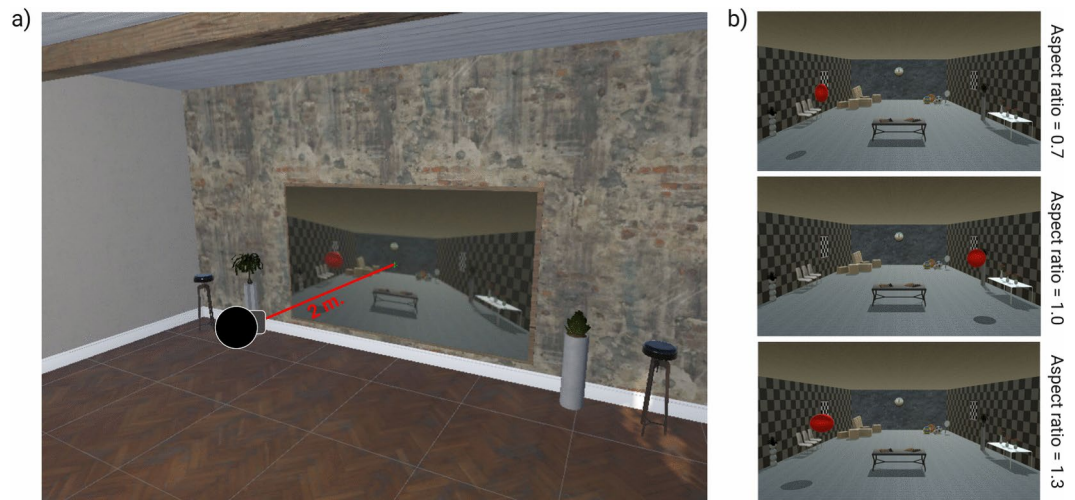


Fig. 10. Virtual environment used during Experiment 3 and Experiment 4. **(a)** In the HMD, participants viewed an environment depicting a living room with realistic furniture, and a large picture display on the wall at a distance of 2 meters, meant to simulate real-life picture viewing, with vergence remaining inside the VR comfort zone⁴⁰. The screen covered most of the participant's field of view. **(b)** In Experiment 4, each picture depicted on the virtual screen from **(a)** showed a realistic scene with some furniture and a red ellipsoid. Viewers were instructed to fixate on a specific location (indicated by a green cross) and to judge whether the ellipsoid was wider or narrower than a perfect sphere. Viewing slant and eccentricity were varied across the experiments.

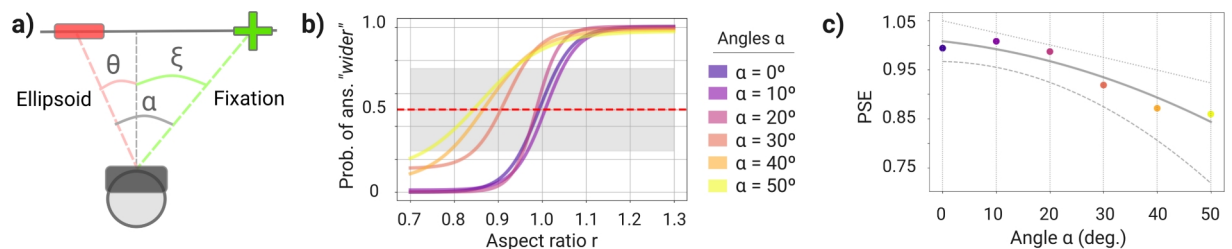


Fig. 11. Overview and results from Experiment 4. **(a)** We showed the participants an image depicting a realistic scene with a salient, red ellipsoid (see Fig. 10) with a slant θ w.r.t. the viewer. We required participants to fixate on a point with eccentricity ξ . **(b)** We fitted one psychometric curve per angle α between the target and the fixation, for six different angles (color-coded). The closer the fixation is to the target (i.e., the smaller the angle), the more accurately people detect the target's width (and the closer the PSE is to 1.0). Larger angles (yellow-ish colored curves) present a clear overestimation of the target's width (i.e., curves are shifted towards the left) **(c)** Point of subjective equality (PSE) for each angle α (i.e., the width of an ellipsoid that would be perceived as a sphere), together with the lower (25%) and upper (75%) detection thresholds in dotted lines. As in **(b)**, for smaller angles (i.e., lower α), people accurately estimate the target being perfectly spherical, yet for larger angles they overestimate its width (e.g., for $\alpha = 50^\circ$, viewers perceive an ellipsoid of width 0.85 as being perfectly spherical).

of fixation eccentricity ξ and slant θ , we varied the projected (image space) aspect ratio r of the ellipsoid. Then, for each possible value α of the visual angle between the ellipsoid and the fixation ($\alpha = \xi - \theta$), we fit a psychometric function to the responses (Fig. 11b), of the form $\psi(r) = 1/(1 + e^{u(r-v)})$ where u and v are curve parameters^{41,42}, and $\psi(r)$ predicts the probability of a participant responding “too wide”. The aspect ratio at which the participants responded at chance level (50%, or $\psi(r; u, v) = 0.5$) is the point of subjective equality (PSE). At the PSE, participants perceive the ellipsoid as spherical. According to the Direct View Condition (DVC), this should happen at the projected aspect ratio $r = 1$. We additionally plotted the PSE as a function of the visual angle α between the ellipsoid and the fixation, Fig. 11c.

For foveal vision and nearby peripheral vision ($\alpha \leq 20^\circ$), the PSE lies close to $r = 1$, indicating that viewers are effectively compensating for local slant. Further away in peripheral vision ($\alpha \geq 30^\circ$), the PSE falls as α increases, indicating that the sphere was perceived as increasingly wider than its image-space aspect ratio r .

What might account for the shape widening in peripheral vision? Lossy peripheral encodings must surely impede shape perception. Yet, Baldwin et al.⁴³ report that real-world objects appear narrower in peripheral vision, whereas we observed the opposite in our experimental setup. Hertzmann¹⁹ hypothesized Fixation

Compensation, in which slanted pictures are perceived as though from a virtual viewpoint with no slant. This also predicts a narrowing of spheres seen in peripheral vision, whereas we observed the opposite. Lossy peripheral encodings could also weaken cues to surface slant, and Vishwanath et al.²² found that compensation decreases with poorer slant cues; this would predict a narrowing of shape. Further experiments are needed to disentangle the effects of peripheral encodings and slant cues.

Regardless, these results suggest that vision accurately perceives shape under surface slant in foveal vision and in peripheral vision near the fixation, but with decreasing accuracy further away from the fixation. In foveal vision, the compensation corresponds to the shape prediction of the DVC. The fact that these peripheral misperceptions do not seem to affect a viewer's experience of pictorial shape further supports the locality of shape perception around fixations.

General discussion

In one telling, viewers understand the real-world by reconstructing a 3D mental representation of the world around them. However, numerous studies have shown inconsistencies in peoples' 3D reasoning, suggesting instead that mental representations are fragmentary⁴⁴, perhaps across different depth ranges or eye fixations.

Likewise, in a conventional understanding, viewers understand a picture as a whole, mentally reconstructing its 3D contents, perhaps by estimating the COP²⁴ and inverting the camera projection⁴⁵. Our Experiment 1 and Experiment 2 contradict conventional picture perception theories based on these ideas, since they show that viewers prefer that depicted objects have their own local linear perspective, rather than being consistent to a projection used for the rest of the picture. This observation aligns with the effectiveness of multiperspective techniques in art history^{4,18} and computational photography^{5,7,11,13}.

Instead, our results support the idea that picture perception operates by per-fixation perspective¹⁹. A viewer interprets detailed shape primarily around the foveal region—sometimes entirely oblivious to shape details in peripheral vision (Experiment 3)—while compensating for surface slant at each fixation (Experiment 4), thereby perceiving shape according to a locally-linear perspective centered at the fixation (Experiment 1 and Experiment 2).

These results explain the perception of shape across different types of projections. In wide-angle linear perspective (Fig. 1(top-left)), the Direct View Condition describes how depicted faces and objects can appear distorted in the margins, where they do not appear as would at the picture center. In contrast, parallel projections (Fig. 1(top-center)) are normally drawn with depicted objects that appear undistorted everywhere. However, all such objects are drawn the same size, without decreasing size as a function of distance to some viewer. In curvilinear perspectives (Fig. 1), the Direct View Condition describes the way that straight lines becoming curved appear distorted, as do other shapes that undergo nonlinear projections. In pushbroom panoramas (Fig. 1(bottom)), different sections may appear similar to linear projections, and appear distorted when they do not. Each of these distortion properties is local to the appearance of pictured objects and shapes within per-fixation perspectives, rather than a global per-picture perspective. The DVC indeed follows in part from the nature of foveal vision. Since viewers do not know how an unfamiliar picture has been constructed, and usually they can only fixate on a small fraction of it at a time, they cannot perceive such a picture according to a global projection. Hence, it is natural to interpret each local image region (around the fixation) separately (Experiment 3). Indeed, interpreting according to a single, global projection (such as a global linear perspective) would additionally require consistency between many successive eye fixations in order to infer the projection, and mentally storing entire pictures at once, which is entirely inconsistent with current evidence (and our own Experiment 3).

Within this broad outline, numerous questions remain. One question is: how much 3D information do viewers obtain from peripheral vision? Baldwin et al.⁴³ report distorted shape judgments in peripheral vision. Our third experiment shows examples in normal viewing where viewers seem entirely oblivious to shape in peripheral vision, whereas our fourth experiment shows that viewers can make some shape judgments in peripheral vision. The differences between these two cases could be explained by many factors, including crowding, memory, and attention, and the specific choices of shapes, projections, and eccentricities. Our fourth experiment shows a dependence of slant compensation on fixation, but how this compensation varies for different stimuli and visual angles requires further study. For example, while participants were able to judge shape in peripheral vision for the stimuli we used, it seems unlikely that they would have done so in normal viewing, and would have only inferred shape within some small visual angle of the fovea. It remains unclear within what ranges of visual angle near the fovea fine-scale shape inference happens during normal viewing.

A broader question, for both real-world vision and pictures, is: how much 3D information do viewers preserve across fixations? We must have some mental 3D representation of the world, but it is far less precise and consistent than commonly assumed⁴⁴. It seems likely that the answers for pictures and real-world vision will share much in common.

Answers to these questions could help explain the various approaches to perspective in art history and aid in the development of new computational photography tools.

Methods

First experiment—stimuli

Participants viewed pairs of images depicting an urban scene containing a set of buildings and street props. We use six commonly-used parametric projections, each of which defines a projection from 3D points to 2D image space in terms of a fixed number of parameters (Fig. 4):

- *Wide-angle linear perspective*, using a horizontal field of view of 60°, approximately equivalent to 31mm focal length.

- *Ultrawide linear perspective*, using a horizontal field of view of 140° , approximately equivalent to 8mm focal length. Within this projection, elements closer to the borders of the image exhibit significant marginal distortion. This is a wider view than smartphone “ultrawide” modes, which are typically 10mm.
- *Isometric projection*³, oriented to view the scene from above.
- *Pushbroom projection*^{5,8,46}, equivalent to capturing the central one-pixel column of a linear perspective camera, while translating the camera horizontally, and then concatenating these pixel columns horizontally. Such projections, also known as “slit-scan,” have been used for visualizing long scenes like street scenes. This projection has an arbitrary horizontal/vertical scale factor caused by the camera’s focal length, its motion relative to the scene, and distance to objects. For a spherical object (like our ball in Fig. 4) captured with a moving camera, different parts of the ball will be recorded at slightly different positions, which leads to horizontal distortions.
- *Equirectangular projection*⁴⁷, created by rendering a 360° panorama and cropping to show only the center portion, including roughly half the picture area.
- *Stereographic projection*, creating the “Little Planet” effect, which is sometimes used for artistic purposes in photography and video^{48,49}.

We used one of two familiar objects in each rendering: a volleyball and a bus. The volleyball is spherical, making it easy to describe its preferred appearances, e.g., the DVC specifies that it should appear circular to look undistorted. The bus was chosen to fit with the urban scene. We used texture-mapped 3D models, so that distortion would be apparent in the texture. For five of the six projections, we generated two different images: the first with the familiar object rendered with the same projection as the rest of the scene (*globally-consistent* condition), and the second with it rendered with a local-linear perspective (*local-linear* condition), according to the DVC. We additionally included a sixth case as a sentinel, i.e., trials for which the expected answer is evident, thus helping us assess that participants were properly focusing on the specified task. For sentinel cases, we rendered one wide-angle projection with minimum marginal distortions with the target objects within the same projection as the scene (which yielded no visible distortions) and another with the targets rendered using ultrawide angle projection (i.e., being significantly distorted, therefore inconsistent with the scene and not satisfying the DVC, see last row of Fig. 4). In these cases, we expected participants to always choose the one with no visible distortions, which besides was the only one congruent with DVC. All the images were rendered with an aspect ratio of 16 : 9, except for the stereographic ones, which were rendered with a 1 : 1 ratio.

Thus, our stimuli consist of *six projections* \times *two target objects* = *twelve pairs of images*. We showed each pair twice per participant, leading to a total of *twenty-four* trials per participant, which were displayed in a randomized order to avoid biases.

First experiment: participants

Twenty-five participants (twelve identified themselves as female, thirteen identified as male, and none of them in other categories or preferred not to say; average age 31.56 years old, $STD = 14.38$) voluntarily took part in the study and provided consent for participation. They were naïve about the final purpose of the experiment.

First experiment: procedure

Participants conducted the experiment using their own computers. The experimental pipeline was designed using Unity 2020.3.32f1. We followed a within-subjects experimental design: Each participant completed a total of twenty-four trials, corresponding to two trials per each of the different twelve conditions. In each trial, the participant was presented with both images side-by-side (i.e., one with the stimulus rendered following the same projection as the rest of the scene, and one with the stimulus rendered with normal projection), and was given unlimited time to answer the question “Which object looks more like a bus?” or “Which object looks more like a volleyball?”. The position of both images on the screen was also randomized. Each response was logged together with the response time. Before starting the experiment, participants were presented with written instructions on how to complete the experiment. Completing the whole set of trials took between 2 and 7 minutes per participant.

Second experiment: stimuli

In the first part of the experiment, and similar to the previous one, participants viewed pairs of images depicting an urban scene containing a set of buildings and street props. However, in this experiment, we focus on linear perspective and evaluate different levels of distortion by adjusting the field of view *FOV* of the camera (Fig. 6). We render the scene with nine different *FOV* $\in \{60^\circ, 70^\circ, \dots, 130^\circ, 140^\circ\}$. Note that 60° and 140° correspond to the previous experiment *wide-angle linear perspective* and *ultrawide linear perspective* respectively, with the remaining serving as intermediate distortion levels between them. As we increased *FOV*, we adjusted the camera position (i.e., put it closer to the props) to capture a similar region of the scene.

As for the previous experiment, we generated pairs of images: one with the yellow bus rendered with the same projection as the rest of the scene (*globally-consistent* condition), and the other one with the bus rendered with a local-linear perspective (*local-linear* condition). Particularly, for each *FOV*, we rendered two of such pairs, each showing the bus in a different orientation (i.e., 225° and 315°), yielding a total of *nine FOV* \times *two orientations* = *eighteen pairs of images*. We showed each pair three times per participant, leading to a total of *fifty-four* trials per participant, which were displayed in a randomized order to avoid biases.

In a second part of the experiment, we asked participants to individually rate the level of distortion of each of the thirty-six rendered images (*nine FOV* \times *two orientations* \times *two conditions*) using a 1–5 Likert scale. Similar to the previous task, we randomized the showing order to avoid biases.

Second experiment: participants

Twenty participants (six identified themselves as female, fourteen identified as male, and none of them in other categories or preferred not to say; average age 26.81 years old, $STD = 5.32$) voluntarily took part in the study and provided consent for participation. They were naïve about the final purpose of the experiment.

Second experiment: procedure

As in the previous experiment, participants conducted the experiment using their own computers. The experimental pipeline was designed using Unity 2020.3.32f1. We followed a within-subjects experimental design. First, each participant completed a total of fifty-four trials, corresponding to three trials per each of the different eighteen pairs of images aforementioned. In each trial, the participant was presented with both images (*globally-consistent* and *local-linear*) side-by-side, and was given unlimited time to answer the question “Which object looks more like a bus?”. The position of both images on the screen was also randomized. After completing the task, each participant completed a total of thirty-six trials in which they were presented each of the rendered images (both *globally-consistent* and *local-linear*) alone, and were given unlimited time to judge the absolute level of distortion of the bus by answering the question “In a scale of 1 (not distorted) and 5 (very distorted), how distorted do you think the yellow bus is?”. Every response was logged together with the response time. Before starting the experiment, participants were presented with written instructions on how to complete the experiment, which took between 12 and 17 minutes per participant.

Third experiment: stimuli

We created a virtual environment comprising a realistic living room, with a large picture display on the front wall, to simulate normal viewing of a screen or artwork on a wall (Fig. 10a). In the virtual environment, participants viewed the display from a distance of 2 meters, so that vergence remained within the comfort zone for our headset⁴⁰, and VR distance underestimation^{50,51} would not play a significant role in our results. During the experiment, we showed different pictures on the display. The *initial condition* picture shows an urban scene, projected with a wide-angle linear perspective camera, with a horizontal field of view of 104° (Fig. 8(top-left)). We created four variants of this picture: in each, an object in one of the four quadrants (i.e., in the side-most 25% of the picture, both horizontally and vertically) is rendered with local-linear perspective (see Fig. 8(right)). The depicted objects were rendered so that they would be positioned in the same region of the picture. Additionally, in order to prevent participants from deducing the type of modifications we were investigating, we added four additional images with other changes, including changing objects’ color, and size, adding an object, or changing one building to one significantly different. (see Fig. 8(bottom-left) for one example). This leads to a total of eight different modified images, in addition to the initial image. All the images were rendered with a 16 : 9 ratio. The experiment consisted of eight different trials, each showing one of the eight aforementioned images.

Third experiment: hardware

We performed the experiment using a Meta Quest Pro head-mounted display (HMD) with a horizontal field of view of 106°, a vertical field of view of 96°, a resolution of 1800x1920 pixels per eye, and a frame rate of 90 fps. Participants could answer the questions using the joystick in a Meta Quest controller. The Meta Quest Pro is provided with inside-out tracking and integrated eye tracking, allowing us to log head and gaze data.

Third experiment: participants

Six participants (one identified herself as female, five identified themselves as male, and none of them identified as non-binary, other, or preferred not to say; average age 28 years old, $STD = 3.16$) voluntarily took part in the study and provided consent for participation. They were naïve about the final purpose of the experiment.

Third experiment: procedure

Participants sat on a non-rotating chair and were instructed not to lean forward or backward during the experiment, to keep them at a fixed distance of the virtual display. The experimental procedure was all designed in Unity 2020.3.32f1.

We followed a within-subjects experimental design, in which all participants got to see the *eight* modified images. The order of the trials was completely randomized to avoid habituation effects. Before starting the actual experiment, we introduced the upcoming procedure to participants, carefully explaining to them the procedure for each trial. Prior to each trial, the virtual display showed an empty, gray background, and the user was indicated which direction (i.e., which of the four quadrants) to look at. After that, the *initial* image appeared with a red cross in the center of the image. After a short time, the red cross blinked for two seconds, and disappeared. Participants were instructed to keep looking at the center of the cross until it completely disappeared, and then direct their gaze toward the direction that had been previously specified. During that period, we tracked their gaze position in real-time, ensuring that it fell in the region indicated by the red cross. If they got out of the region containing the red cross for more than 200 ms (note that an average human saccade usually lasts 100 ms^{52,53}), then the trial was marked as invalid.

When the eye tracker detected that their gaze left the central region and entered the corresponding quadrant of the image, the display blanked for 250 ms^{37,39}, which is in the range of an average human blink (100 to 400 ms)^{52,54}. We artificially simulated such blinks since we did not want to explicitly ask participants to blink (which could distract them from the main task), nor could we reliably detect blinks with the integrated eye tracker (i.e., it is not possible to disambiguate whether the lack of gaze information is because of a blink or missed data). During the blink, we changed the depicted scene to one with a modification in the corresponding quadrant. When the blink ended, the modified image was already displayed. If the participant had not moved their gaze when the blink had ended, the trial was also marked as invalid. Then, we gave the participants another two seconds to

look at that particular quadrant, and removed the image, showing again a gray background. If the participants directed their gaze toward an incorrect quadrant, the trial would also be marked as invalid. We finally asked the participants whether they had detected any change between the scene before and after the blink. Note that our approach is conservative by definition since participants are explicitly aware that changes may happen (in contrast to a natural scenario where they would be naïve, therefore being even less sensitive to changes^{37,55}).

After the trials, we showed participants two pictures, and asked which of them they had seen at some time during the experiment. The images were the *initial condition* image, which was shown at the start of each trial, and a *composite* image with all of the perspective modifications seen during trials (Fig. 8(bottom-left)). Specifically, each quadrant in the composite image contained the perspective modification that the viewer had seen while fixating on that quadrant in a trial; this image had never been shown during the trials, but we hypothesized that viewers might not realize they had not seen it.

Fourth experiment: stimuli

We used the same virtual environment as in the third experiment. Unlike it, in this experiment, the stimulus within the displayed picture consisted of rendered images of a room, containing several props for context, and a red ellipsoid (Fig. 10). The picture surface was only slanted horizontally with respect to the viewer, and thus only horizontal foreshortening appeared. Hence, foreshortening would have the effect of horizontal compression of the shape in the retinal image.

We rendered ellipses with aspect ratios $r \in \{0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3\}$, where r is the image-space width of the ellipsoid divided by its height. Ellipsoids had the same width in the forward- and left-axis (i.e., they always looked like they were facing the viewer) and were rendered in two different positions, so that, when displayed in the virtual environment, they would be at a slant w.r.t. the viewer's position of $\theta = \pm 20^\circ$, leading to a total of 14 different images. During the experiment, participants were asked to fixate at six different eccentricities ξ to create six different angles α between the ellipsoid and the fixation such that $\alpha = \xi - \theta \in \{0^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ, 50^\circ\}$. Each of the aforementioned 14 images was seen from each of those angles twice, therefore yielding 168 trials per participant. This experimental setup allows us to compute one psychometric curve per angle α .

Fourth experiment: hardware

We use the same hardware as in the third experiment.

Fourth experiment: participants

Six participants (two identified themselves as female, four identified themselves as male; average age 28.5 (STD = 1.60)) voluntarily took part in the study and provided consent for participation. They all were naïve about the final purpose of the experiment.

Fourth experiment: procedure

Participants sat on a non-rotating chair and were instructed not to lean forward or backward during the experiment, to keep them at a fixed distance of the virtual display. The experimental software was developed in Unity 2023.2.20f. We followed a within-subjects experimental design, in which all participants got to see all 168 trials, with randomized presentation order. Participants were presented with a brief tutorial introducing them to the procedure and then proceeded with the experiment itself. Each trial began with a gray background displayed on the screen. After half a second, a green cross appeared at the fixation eccentricity ξ corresponding to that trial. Participants were instructed to keep looking at the green cross all the time; which we assessed using the HMD integrated eye-tracker. After another half a second, the gray background was replaced by the corresponding rendered image, containing the ellipsoid at the corresponding slant θ , which remained on the display for one second. After that, the displayed image and the green cross disappeared, and participants were asked to answer the question “Was the sphere too wide or too narrow?” The whole procedure took around twenty minutes per participant, and we allowed them to take a break at any moment to avoid fatigue. None of them requested it.

We conducted this experiment as a two-alternative forced-choice (2AFC) task with a method of constant stimuli (the shape and slant of the ellipsoid and the fixation eccentricity did not depend on the trial order, but were distributed randomly). This is a common procedure to find and analyze psychophysical phenomena^{41,56}, which produces more robust results than directly asking for absolute judgments. Since participants are forced to choose one of the two options, their answer will be random (close to chance level) for ellipsoids whose aspect ratio is perceived to be close to 1.0 (i.e., they are perceived as spheres). Aggregating measurements from answers from different ellipsoid shapes and eccentricities allowed us to fit psychometric curves modeling participants' shape perception.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request. Point of contact: danims@unizar.es.

Received: 22 August 2024; Accepted: 22 April 2025

Published online: 02 May 2025

References

1. Gombrich, E. H. *Art and Illusion: A Study in the Psychology of Pictorial Representation* (Princeton University Press, 1961), 2nd edn.
2. Kemp, M. *The Science of Art: Optical themes in western art from Brunelleschi to Seurat* (Yale University Press, 1990).
3. Willats, J. *Art and Representation: New Principles in the Analysis of Pictures* (Princeton University Press, 1997).

4. Agrawala, M., Zorin, D. & Munzner, T. Artistic multiprojection rendering. In *Eurographics Workshop on Rendering Techniques* 125–136 (Springer, 2000).
5. Román, A., Garg, G. & Levoy, M. Interactive design of multi-perspective images for visualizing urban landscapes. In *IEEE Visualization 2004* 537–544 (IEEE, 2004).
6. Zelnik-Manor, L., Peters, G. & Perona, P. Squaring the circle in panoramas. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2* 1292–1299 (IEEE, 2005).
7. Agarwala, A., Agrawala, M., Cohen, M., Salesin, D. & Szeliski, R. Photographing long scenes with multi-viewpoint panoramas. *ACM Trans. Graphics* **25**, 853–861 (2006).
8. Román, A. & Lensch, H. P. Automatic multiperspective images. *Render. Techn.* **2**, 161–171 (2006).
9. Zelnik-Manor, L. & Perona, P. Automating joiners. In *Proceedings of the 5th International Symposium on Non-Photorealistic Animation and Rendering, NPAR '07* 121–131. <https://doi.org/10.1145/1274871.1274890> (Association for Computing Machinery, 2007).
10. Carroll, R., Agrawala, M. & Agarwala, A. Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph.* **28**, 23 (2009).
11. Badki, A., Gallo, O., Kautz, J. & Sen, P. Computational zoom: a framework for post-capture image composition. *ACM Trans. Graph.* **36**, 56 (2017).
12. Shih, Y., Lai, W.-S. & Liang, C.-K. Distortion-free wide-angle portraits on camera phones. *ACM Trans. Graph.* **38**, 1–12 (2019).
13. Liu, S. J., Agrawala, M., DiVerdi, S. & Hertzmann, A. Zoomshop: depth-aware editing of photographic composition. *Comput. Graph. Forum* **41**, 57–70 (2022).
14. Elkins, J. *The Poetics of Perspective* (Cornell University Press, 1994).
15. Verstegen, I. A classification of perceptual corrections of perspective distortions in renaissance painting. *Perception* **39**, 677–694 (2010).
16. Kubovy, M. *The Psychology of Perspective and Renaissance Art* (Cambridge University Press, 1986).
17. Koenderink, J., van Doorn, A., Pinna, B. & Pepperell, R. On right and wrong drawings. *Art and Perception* **4**, 1–38 (2016).
18. Perona, P. Far and yet close: multiple viewpoints for the perfect portrait. *Art & Perception* **1**, 105–120 (2013).
19. Hertzmann, A. Toward a theory of perspective perception in pictures. *J. Vis.* **24**, 896 (2024).
20. Sharpless, T. K., Postle, B. & German, D. M. Pannini: a new projection for rendering wide angle perspective images. In *Computational Aesthetics* 9–16 (2010).
21. Pepperell, R. & Haertel, M. Do artists use linear perspective to depict visual space?. *Perception* **43**, 395–416 (2014).
22. Vishwanath, D., Girshick, A. R. & Banks, M. S. Why pictures look right when viewed from the wrong place. *Nat. Neurosci.* **8**, 1401–1410 (2005).
23. Cooper, E. A., Piazza, E. A. & Banks, M. S. The perceptual basis of common photographic practice. *J. Vis.* **12**, 8–8 (2012).
24. Pirenne, M. H. *Optics, Painting & Photography* (Cambridge University Press, 1970).
25. 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 (1976).
26. Rosinski, R. R. & Farber, J. Compensation for viewing point in the perception of pictured space. In *The Perception of Pictures: vol. I. Alberti's Window: The Projective Model of Pictorial Information* (ed. Hagen, M.) (Academic Press, 1980).
27. Adams, K. R. Perspective and the viewpoint. *Leonardo* **5**, 209–217 (1972).
28. Bengston, J. K., Stergios, J. C., Ward, J. L. & Jester, R. E. Optic array determinants of apparent distance and size in pictures. *J. Exp. Psychol. Hum. Percept. Perform.* **6**, 751 (1980).
29. Todorović, D. Is pictorial perception robust? The effect of the observer vantage point on the perceived depth structure of linear-perspective images. *Perception* **37**, 106–125 (2008).
30. Zorin, D. & Barr, A. H. Correction of geometric perceptual distortions in pictures. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* 257–264 (1995).
31. Rosenholtz, R. Capabilities and limitations of peripheral vision. *Annu. Rev. Vis. Sci.* **2**, 437–457 (2016).
32. Rosenholtz, R. Demystifying visual awareness: peripheral encoding plus limited decision complexity resolve the paradox of rich visual experience and curious perceptual failures. *Atten. Percept. Psychophys.* (2020).
33. Barre, A. & Flocon, A. *La perspective curviligne: de l'espace visuel à l'image construite* (FeniXX, 1968).
34. Ward, J. L. The perception of pictorial space in perspective pictures. *Leonardo* **9**, 279–288 (1976).
35. Pepperell, R., Ruta, N. & Burleigh, A. Egocentric vision in a 3d game using linear perspective and natural rendering. In *Proc. EGOAPP* (2019).
36. Freeman, J. & Simoncelli, E. P. Metamers of the ventral stream. *Nat. Neurosci.* **14**, 1195–1201 (2011).
37. Martin, D., Sun, X., Gutierrez, D. & Masia, B. A study of change blindness in immersive environments. *IEEE Trans. Visualiz. Comput. Graph.* **29**, 2446–2455 (2023).
38. Hollingworth, A. Object-position binding in visual memory for natural scenes and object arrays. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 31 (2007).
39. Ma, L.-Q., Xu, K., Wong, T.-T., Jiang, B.-Y. & Hu, S.-M. Change blindness images. *IEEE Trans. Visualiz. Comput. Graph.* **19**, 1808–1819 (2013).
40. Shibata, T., Kim, J., Hoffman, D. M. & Banks, M. S. The zone of comfort: predicting visual discomfort with stereo displays. *J. Vis.* **11**, 11–11 (2011).
41. Serrano, A., Martin, D., Gutierrez, D., Myszkowski, K. & Masia, B. Imperceptible manipulation of lateral camera motion for improved virtual reality applications. *ACM Trans. Graph.* **39**, 69 (2020).
42. Steinicke, F., Bruder, G., Jerald, J., Frenz, H. & Lappe, M. Estimation of detection thresholds for redirected walking techniques. *IEEE Trans. Visualiz. Comput. Graph.* **16**, 17–27 (2009).
43. Baldwin, J., Burleigh, A., Pepperell, R. & Ruta, N. The perceived size and shape of objects in peripheral vision. *i-Perception* **7**, 56 (2016).
44. Linton, P. et al. New approaches to 3D vision. *Phil. Trans. R. Soc. B* **378**, 145 (2022).
45. Juricevic, I. & Kennedy, J. Looking at perspective pictures from too far, too close, and just right. *J. Exp. Psychol. Gener.* **135**, 448–61 (2006).
46. Seitz, S. M. & Kim, J. Multiperspective imaging. *IEEE CG & A* **23**, 56 (2003).
47. Sitzmann, V. et al. Saliency in vr: how do people explore virtual environments?. *IEEE Trans. Visualiz. Comput. Graph.* **24**, 1633–1642 (2018).
48. Swart, D. & Torrence, B. Mathematics meets photography part i. *Math Horizons* **19**, 14–17 (2011).
49. German, D. M. et al. Flattening the viewable sphere. In *CAE* 23–28 (2007).
50. El Jamiy, F. & Marsh, R. Distance estimation in virtual reality and augmented reality: a survey. In *2019 IEEE International Conference on Electro Information Technology (EIT)* 063–068 (IEEE, 2019).
51. Creem-Regehr, S. H., Stefanucci, J. K. & Bodenheimer, B. Perceiving distance in virtual reality: theoretical insights from contemporary technologies. *Philos. Trans. R. Soc. B* **378**, 20210456 (2023).
52. Nyström, M. & Holmqvist, K. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behav. Res. Methods* **42**, 188–204 (2010).
53. Baloh, R. W., Sills, A. W., Kumley, W. E. & Honrubia, V. Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology* **25**, 1065–1065 (1975).

54. Salthouse, T. A. & Ellis, C. L. Determinants of eye-fixation duration. *Am. J. Psychol.* **1980**, 207–234 (1980).
55. Suma, E. A. et al. Leveraging change blindness for redirection in virtual environments. In *Proc. IEEE VR* 159–166 (IEEE, 2011).
56. Langbehn, E., Steinicke, F., Lappe, M., Welch, G. F. & Bruder, G. In the blink of an eye: leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Trans. Graph. (TOG)* **37**, 1–11 (2018).

Acknowledgements

Part of this work was conducted when Daniel Martin was an intern at Adobe Research, and we would like to acknowledge Adobe's support. This work has also been supported by grant PID2022-141539NB-I00, funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU. This project has been partially funded by the Government of Aragon's Departamento de Educación, Ciencia y Universidades through the project "HUMAN-VR: Development of a Computational Model for Virtual Reality Perception" (PROY_T25_24). Daniel Martin was supported by a Gobierno de Aragon predoctoral grant (period 2020-2024). The authors would like to thank Ana Serrano and Julia Guerrero-Viu for insightful discussions, and Ruth Rosenholtz and Dawn Wilson for their comments and suggestions on the manuscript.

Author contributions

D.M., S.D., A.H. designed the research; D.M. performed the research; D.M., D.G., B.M., S.D., A.H. analyzed the data; D.M., D.G., B.M., S.D., A.H. wrote the paper.

Competing interests

The authors declare no competing interests.

Ethics statement

The experiments were conducted in accordance with the guidelines and regulations of Universidad de Zaragoza (Spain). Our experimental protocols comply with the requirements approved by the Consejo de Gobierno (Government Council) of Universidad de Zaragoza. Written informed consent was obtained from participants before experiments began, and particular attention was paid to ensure that research data could be curated in an anonymized manner. At the outset of the experiment, it was made clear to participants that they participated voluntarily and that they had the right to withdraw from the research at any time without giving a reason. Subjects appearing in the images of this manuscript have given written informed consent for publication of identifying information or images in an online open-access publication.

Additional information

Correspondence and requests for materials should be addressed to D.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025