Mirror-reflecting a picture of an object: What happens to the shape percept?

ELS V. K. CORNELIS, ANDREA J. VAN DOORN, and HUIB DE RIDDER Delft University of Technology, Delft, The Netherlands

Although a two-dimensional picture never fully specifies the actual layout of the depicted three-dimensional scene, one is still able to make a three-dimensional interpretation. When a picture is mirror-reflected, the range of plausible scenes possibly corresponding with the depicted scene has not changed with respect to the original depiction. We were curious to find out whether the inherent picture ambiguities would be solved the same way or differently. Participants performed local attitude settings on three sets of pictures: (1) original pictures, (2) left-right mirrored pictures, and (3) up-down mirrored pictures. Pairwise comparison of the pictorial reliefs of the depicted object, reconstructed from the raw settings, revealed dissimilarities. The differences, however, could be drastically diminished by conducting an affine transformation correction taking into account not only the depths, but also the picture plane coordinates. The inherent ambiguities seemed thus to be solved differently between conditions. By factoring out different solutions to the ambiguities, the pictorial reliefs were found to be equivalent.

Looking around in the world, we are encountered by all kinds of objects presented to us in all kinds of environmental circumstances. For instance, it can be a bright sunny day. But it can be dark and cloudy as well. Although the transient properties of objects may change drastically under different environmental circumstances, object recognition is usually fairly easy.

An example, from daily life, of nearly continuously changing viewing (environmental) conditions is the variation in viewpoint. The projection of the shape of the object on the retina is altering all the time. Yet in many cases, the percept of the object remains the same. This phenomenon is called *shape constancy*.

On Objects and Viewpoint

In previous research, shape perception across viewpoints has been a topic of investigation. When two-dimensional (2-D) figures were used, shape constancy was found to be rather accurate, although diminishing with greater slant (Thouless, 1931). However, three-dimensional (3-D) objects—wire or clay clumps—seemed to have a much poorer shape constancy (Rock & DiVita, 1987; Rock, DiVita, & Barbeito, 1981).

Even so, in daily life, people moving around usually experience few problems in identifying and handling the surrounding objects correctly and appropriately. Survival would not have been possible otherwise.

Two major classes of theories accounting for object recognition across viewpoints can be distinguished, de-

Correspondence concerning this article should be addressed to E. V. K. Cornelis, Industrial Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands (e-mail: e.v.k.cornelis@io.tudelft.nl).

pending on the importance of the viewpoint of the observer. In addition, in a third class of theories, there is a compromise between the two views.

The first class of accounts stresses that visual recognition is viewpoint dependent. In this case, the observed view of the object should be matched against a template in order to recognize it. A one-to-one matching process with as many representations stored as there are views of an object is not parsimonious. Storing one specific, "representative" view is. Such a view is called a canonical perspective (Palmer, Rosch, & Chase, 1981). The greater the angle of rotation between the canonical (e.g., previously presented) view of the object and the presented view, the longer it takes to identify the presented object as the same or a different one. The time spent making the same/different decision is thought to be used to make a mental rotation in order to match the presented object with its mental representation (Shepard & Cooper, 1982; Tarr, Bülthoff, Zabinski, & Blanz, 1997).

The second approach denies the influence of viewpoint-dependent recognition, proposing that objects have specific viewpoint-invariant characteristics. In a theoretical paper, Marr and Nishihara (1978) presented an object-centered coordinate system in which each object has a 3-D model representation based on generalized cones. Biederman (e.g., 1987) suggested that objects are composed of geons—that is, simple geometric components, all having clear-cut properties that are viewpoint independent.

A third class of theories takes a more nuanced position in this debate: Both viewpoint-dependent and viewpoint-independent processes possibly occur, depending on the circumstances (e.g., Tarr, 1995; Vanrie, Willems, & Wagemans, 2001; Wagemans, Van Gool, Lamote, & Foster, 2000).

In all the studies mentioned hitherto, a decision paradigm in which visual recognition played a major role was used. Participants were instructed to decide whether a stimulus was the same or different. To this end, one had to draw on recognition processes to compare the presented object with a previously or simultaneously shown one.

Nonetheless, it is also possible to investigate the phenomenon of shape constancy more directly. In this type of study, probing of the percept of the shape does not involve a mediating decision process in which visual recognition is essential. Hence, one does not have to take a roundabout route to compare recognized percepts of an object's shape shown under different conditions. Comparing *decisions* based on the recognized percept of the shape of an object is unnecessary. One can simply compare the probed *percepts* themselves. In the present research, we used this direct approach.

Koenderink, van Doorn, and Kappers (1995), for instance, probed the percepts of the shape of a real object—that is, a sphere—by varying the viewer's characteristics. Participants manipulated a gauge figure projected onto the object's surface until it visually seemed to fit. Comparison of the depth reliefs obtained when the sphere was looked at binocularly, monocularly, and synoptically showed depth scalings that depended on the condition. The depth relief was, respectively, deepest, flatter, and even flatter. Using a similar method, van Doorn (2000) found that scene conditions systematically influenced the perception of the shape of objects: The size of an object, as well as the distance from the vantage point, mattered significantly.

On Pictures and Reflection

In order to fully understand, say, a sculpture, one generally has the tendency to walk around it and look at it from all sides. This way, it is possible to integrate the different parts into a whole.

Pictorial perception, however, is a special case of visual perception. Two different worlds are intertwined. A picture can be crumpled or torn up: It is a physical object, a piece of paper covered with pigments organized in a specified way. Furthermore, a picture also *depicts* a scene one can observe when one looks "into" it (Gibson, 1979; Koenderink & van Doorn, 2003; Pirenne, 1970).

Although one can rotate the physical object (i.e., the piece of paper), a picture always forces the observer to look at the depicted object from a well-specified vantage point—namely, the lens position of the camera. Depicting a 3-D object necessarily results in losing the distancerange. Yet, the observer is still capable of interpreting this object in a 3-D space. The 3-D interpretation of the depicted object is called the *pictorial relief*.

To overcome the limitation of 2-D representations, it is of course possible to depict an object from several viewpoints. Representing an object in different views results in different pictures. In this respect, it is interesting to briefly mention a study in which correspondences were studied between pictures of different views of the same object (Koenderink, Kappers, Pollick, & Kawato, 1997).

Each time, two pictures, showing different views of the object, were presented side by side on a computer screen. In one picture, a dot appeared that was located on the image of the object. Participants were asked to indicate the same location on the rotated object depicted in the other picture. It was found that the participants were indeed able to perceptually set these correspondences. More important, the participants' conduct of this task seemed to be founded on their pictorial relief. It is rather remarkable that people can match locations on pictures differing in structural information with such ease, since until now, no one has succeeded in implementing this capacity in machine vision.

Up till now, all studies reporting on the probing of the 3-D interpretation of depicted objects directly have focused on observer viewing conditions (for instance, monocular, binocular, or synoptical viewing; Koenderink, van Doorn, & Kappers, 1994), cue differences in the picture material (for instance, lighting conditions [Koenderink, van Doorn, Christou, & Lappin, 1996a, 1996b; Todd, Koenderink, van Doorn, & Kappers, 1996], or rendering conditions [Koenderink et al., 1996b]). In general, the results of these studies lead to the conclusion that shape constancy is the case throughout, although systematical differences exist that suggest that there is a significant influence of the conditions on the pictorial relief.

In this article, we are interested in the question of to what extent shape perception of a depicted object is symmetrical. Therefore, we will investigate the influence of the reflection of a picture on the 3-D interpretation of the depicted object. If shape perception of depicted objects is symmetrical, the pictorial reliefs obtained from the original picture and its mirrored counterpart will be identical; if shape perception of depicted objects is not symmetrical, we will find differences between the pictorial reliefs obtained from the original picture and its mirrored version.

Thus, in the present experiment, we neither varied the observer's viewing conditions nor manipulated the conditions under which the object was photographed, hence altering the depth information of the picture. The pictorial cues thus were not varied in any way: Mirror-reflecting a picture consisted only of interchanging the pixels around the middle axis. The picture thus was "changed, yet the same" (Stewart & Golubitsky, 1993, p. 28).

EXPERIMENT

In order to study the effect of the reflection of pictures on the 3-D interpretation of the depicted object, we mirrored pictures around the horizontal axis and around the vertical axis. There were thus three conditions: (1) original pictures, (2) left–right mirrored pictures, and (3) up–down mirrored pictures.

The object depicted was a female shop window mannequin. The torso was chosen because of its smooth surface and its geometrical structure. It was highly generic.

Six pictures were taken, each from a different vantage point: three dorsal and three frontal views with pose differences of 30°. This difference brought variation in the

stimulus material, whereas between poses there still was sufficient overlap.

The pictorial relief was externalized using the same method as that developed by Koenderink, van Doorn, and Kappers (1992). Participants performed local attitude settings by adjusting the shape and the orientation of an elliptical gauge figure projected onto the object's surface. The participant's responses—that is, the local attitude settings—reflected the perceived orientation of the local surface. They defined the slant and the tilt values of the local tangent plane. The slant can easily be understood as the angle between the line of sight and the surface normal. The tilt is, in fact, the direction of the slant (Figure 1). The pictorial reliefs were constructed from the local attitude settings.

METHOD

Stimuli

The stimuli consisted of pictures of a female torso used in shop windows to display clothing.

The torso was made out of plastic material and showed diffuse reflection. By means of a turntable provided with a reference direction, the torso could be placed in well-specified poses. The turntable, as well as the camera, had a fixed position in space. In the experiment, the poses differed by 30° around the vertical axis. Three poses were selected for both the frontal (60°, 90°, and 120°) and the dorsal (220°, 250°, and 280°) views. The light source was positioned at a fixed oblique top-left location in space.

The torso was photographed in grayscale, using a Nikon Coolpix 990 digital camera.

The pictures of all the poses were mirrored around the vertical and the horizontal axes. This way, three conditions were obtained: (1) original, (2) left-right mirrored, and (3) up-down mirrored pictures (Figure 2).

Procedure

The photographs were displayed in the center of a 21-in. flat computer screen, type LaCie elektron 22 blue, connected to a Power Mac G4. The size of the picture measured 600×800 screen pixels $(18.6 \times 23.8 \text{ cm})$.

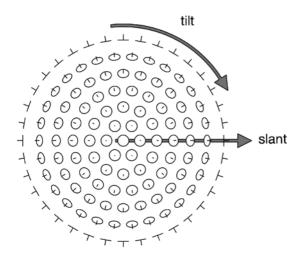


Figure 1. Impression of gauge figure settings. The gauge figure settings define the slant and the tilt values of the local tangent plane.

We used a chinrest in order to limit head movements of the participants. The distance between the viewpoint of the participants and the screen measured 50 cm. The chinrest was mounted so that the center of the screen was straight ahead of the viewpoint.

The room was dimly lit, and the contours of the computer screen were vaguely visible.

The participants were asked to adjust the shape and the orientation of a probe superpositioned onto the surface of the torso in such a way that it would visually fit the torso's surface. The probe was somewhat similar to a thumbtack: It consisted of the outline of a red elliptical gauge figure and a line segment sticking out from the center. The long axis of the ellipse was always 40 pixels; the maximum length of the line segment measured 20 pixels. After being adjusted appropriately, the ellipse had to look like a circle painted on the surface, with the line segment sticking out in the normal direction (Figure 3).

In order to adjust the probe, the participants manipulated a track ball. The participants were instructed to click a button whenever they were satisfied with their setting. The attitude parameters for the setting (slant and tilt value) were then saved, and the next probe appeared at a different position on the picture. During the run of an experiment, only one probe was visible at a time.

The participants performed 483–528 local attitude settings per stimulus, one after another. Each participant ran three sessions per stimulus. In this way, extensive data sets were obtained.

All the pictures were triangulated in the following way. First, the contour of the depicted object was traced onto the picture. Next, a lattice of equilateral triangles was put on the image plane within the previously specified contour. The barycentra of the triangles generated by this triangulation determined the positions of the probe. The positions were presented in a random order. The participants carrying out the experiment were not aware of the triangulation.

The participants executed the task monocularly to improve the impression of depth in the picture (Koenderink et al., 1994).

Participants

The participants were the authors: A.D., E.C., and H.R. All had normal or corrected-to-normal vision. Only A.D. had seen the actual object.

RESULTS

Data Evaluation

As was previously mentioned, the participants altered the shape and the orientation of the ellipse in order to make it look like a circular figure painted on the object's surface. Adjusting the gauge figure was, for all the participants, a natural experience. The participants ran three sessions for each stimulus. Each session consisted of 483–528 local attitude settings, executed sequentially in random order, and took about half an hour, on average. The position of each probe was fully determined by the triangulation of the torso carried out on all the pictures. The gauge figure settings reflected the perceived local attitude of the surface at the barycentra of the triangles. The shape of the ellipse informed us about the local slant, its orientation about the tilt. Slant values measured the angle between the line of sight and the surface normal at the probe's position and ranged from 0° to 90°. Thus, a slant with an angular value of 0° and, therefore, corresponding with a local surface normal coinciding with the line of sight would mean that the local tangent plane was perpendicular to the line of sight. A slant measuring 90° would mean that the local surface normal was directed perpendicularly to the line of sight; hence, the

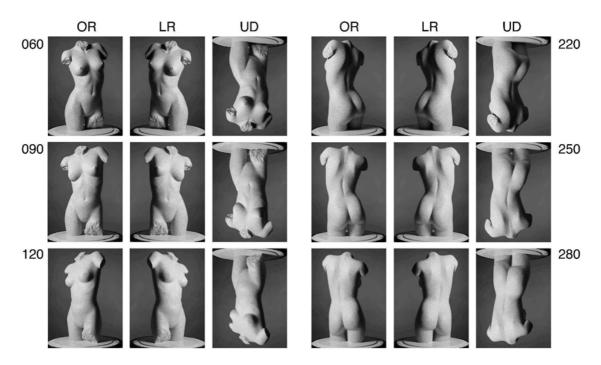


Figure 2. Stimulus material. Six pictures were taken of a female torso: three frontal views $(60^\circ, 90^\circ, \text{and } 120^\circ)$ and three dorsal views $(220^\circ, 250^\circ, \text{and } 280^\circ)$. All six pictures were mirror-reflected around the vertical axis (left-right mirrored) and around the horizontal axis (up-down mirrored). OR, original picture; LR, left-right mirrored picture; UD, up-down mirrored picture.

local tangent plane was parallel to the line of sight. The orientation of the slant, the tilt, included angular values from -180° to 180° (Figure 1).

In order to derive a pictorial relief constructed from the responses of the participant, the following procedure, extensively described in Koenderink et al. (1992), was used. On the basis of the attitude settings performed in each session, gradients were calculated at each probe position. Subsequently, the depth differences over the faces of the triangulation were computed. Next, the mean of

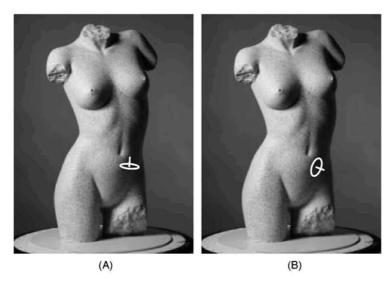


Figure 3. Procedure of the experiment. One probe at a time appeared on the object's surface. The participant adjusted the gauge figure so that it seemed to fit the object's surface. (A) An example of a setting of the gauge figure that did not visually fit the object's surface. (B) An example of a possible setting of the gauge figure that visually fitted the object's surface.

the depth values was set to zero, since a depth offset of the pictorial relief would be irrelevant. Finally, the depth values were determined, providing the best fit with the attitude settings in the least squares sense.

For each participant, the depth lists were averaged over sessions. Finally, the depths obtained for all the mirrored pictures were mirrored back, in order to make a direct comparison between the pictorial reliefs of the original pictures and their manipulated versions feasible. The resulting pictorial reliefs are shown in Figures 4A–4C.

Next, we will attempt to answer our main question: Does mirror-reflecting a picture influence the 3-D interpretation of the depicted object. To this end, we will systematically compare the pictorial reliefs obtained for the original pictures and their (left-right or up-down) mirrored counterparts for each pose. Later on, we will explore interindividual differences between the pictorial reliefs acquired from the different participants.

Comparing Conditions

Externalizing the pictorial relief in this way enables visually examining one's operationalized percept. In investigating the influence of mirror reflection of the picture on the 3-D interpretation of the depicted object, preliminary insight can be gained by visually comparing the constructed pictorial reliefs obtained for the original pictures and for their mirrored counterparts. Figures 4A–4C show the profile views of the reconstructed pictorial reliefs of

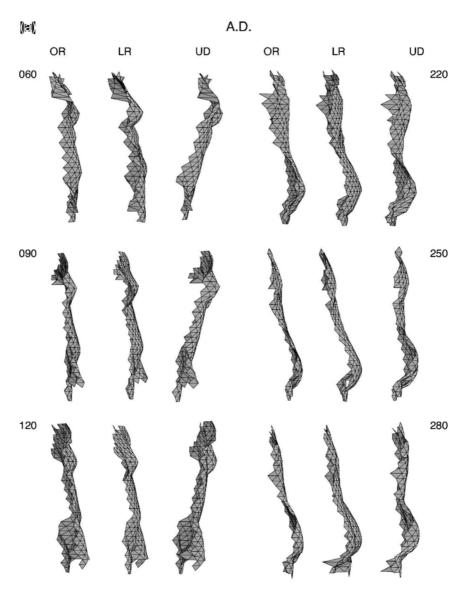


Figure 4A. Profile views of the reconstructed pictorial reliefs of the depicted object obtained from all poses in all conditions for Participant A.D. The depths obtained from the mirrored pictures were mirrored back in order to make a direct comparison between the pictorial reliefs possible. 060, 090, 120, 220, 250, and 280 indicate the various poses in degrees. OR, original picture; LR, left-right mirrored picture; UD, up-down mirrored picture.

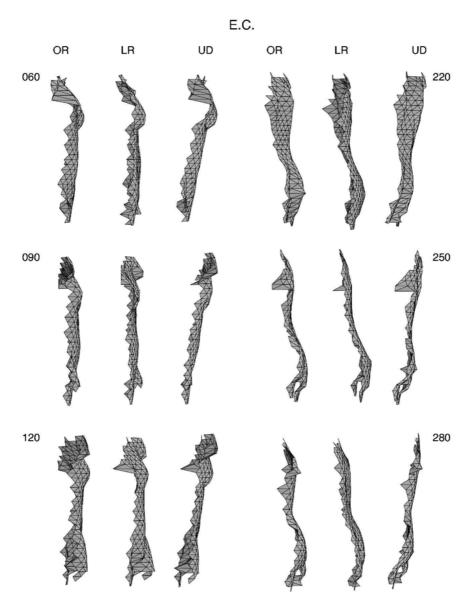


Figure 4B. Profile views of the reconstructed pictorial reliefs of the depicted object obtained from all poses in all conditions for Participant E.C. The depths obtained from the mirrored pictures were mirrored back in order to make a direct comparison between the pictorial reliefs possible. 060, 090, 120, 220, 250, and 280 indicate the various poses in degrees. OR, original picture; LR, left-right mirrored picture; UD, up-down mirrored picture.

the depicted object for all conditions of all poses for each participant. Note that the pictorial reliefs of the original picture and the mirrored versions are not entirely identical. Especially, the up—down pictorial reliefs are strikingly different, as compared with the reliefs of their originals.

In the following, we will quantitatively examine the differences occurring between the reconstructed 3-D interpretations of the depicted torso obtained from the original and the mirrored pictures. First, we will focus on the differences between conditions in a global way. Subsequently, we will elaborate briefly on piecewise differences between conditions.

Henceforth, we will indicate the original picture/left-right mirrored version comparison by *OR-LR*, and the original picture/up-down mirrored version combination by *OR-UD*.

Global Differences

Straight regression. A quantitative analysis of the relation of the depth values between OR–LR and between OR–UD was carried out by performing linear regressions on the depth lists¹ for each pose–participant combination separately. This correlation model showed percentages of explained variance ranging from 0.8% to

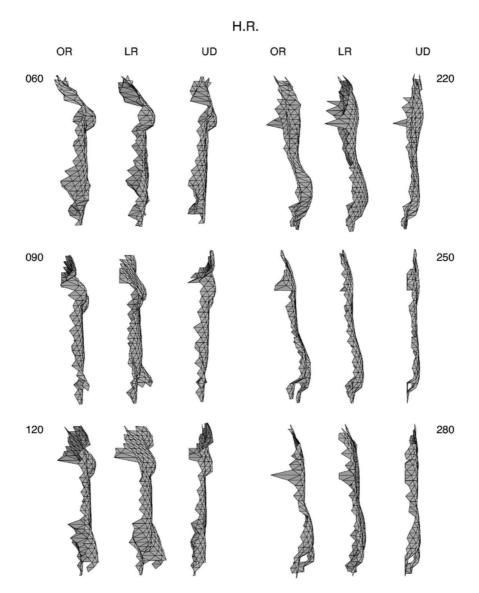


Figure 4C. Profile views of the reconstructed pictorial reliefs of the depicted object obtained from all poses in all conditions for Participant H.R. The depths obtained from the mirrored pictures were mirrored back in order to make a direct comparison between the pictorial reliefs possible. 060, 090, 120, 220, 250, and 280 indicate the various poses in degrees. OR, original picture; LR, left-right mirrored picture; UD, up-down mirrored picture.

98.0% for Participant A.D. For Participant E.C., the depth values of the original picture explained from 1.3% to 93.9% of the variance in the depth values of the mirrored picture. The R^2 s of Participant H.R. varied from 12.5% to 93.3% (see the black bars in Figure 7). All R^2 s were significant: for A.D., all ps < .0001, except that for 060UD, p = .1161, for 090UD, p = .0688, and for 120UD, p = .0006; for E.C., all ps < .0001, except that for 090UD, p = .0019, and for 280UD, p = .0510; for H.R., all ps < .0001. The OR–LR comparison systematically yielded high coefficients of determination. All scatterplots relating the depth values

of the original pictures and their left-right mirrored versions generally displayed rather linear relationships between both depth lists, suggesting that similar reliefs were constructed (Figure 5). The regression model for OR-UD performed much worse (see the black bars in Figure 7). The OR-UD scatterplots showed somewhat chaotic patterns (Figure 5).

Regressing the depth values of the original picture onto the depth values of the left-right or the up-down mirrored picture over all locations indicates to what extent there is a depth offset and a depth scaling between the pictorial reliefs of the two pictures. The intercept re-

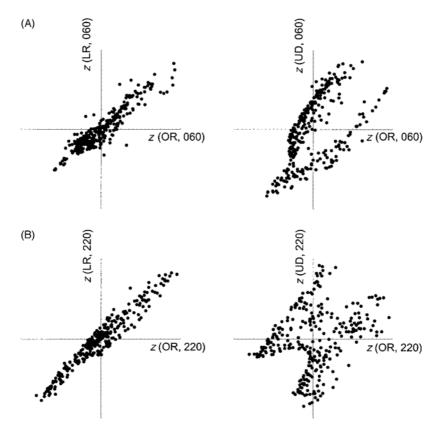


Figure 5. Scatterplots of the raw depth values obtained from the original picture (x-axis) versus the raw depth values obtained from the left–right or the up–down mirrored picture (y-axis) for (A) pose of 60° and (B) pose of 220° , Participant E.C. z, depth values, measured in pixels; OR, original picture; LR, left–right mirrored picture; UD, up–down mirrored picture. The x- and the y-axes range from -100 to 100 pixels.

ferring to the depth offset is irrelevant. The weight of the depth z (OR) reveals the extent and the nature of the depth scaling appearing in the pictorial relief of the mirrored picture. Unit slope differences higher than 1 or lower than 1 indicate, respectively, more or less depth relief in the mirrored picture, as compared with the depth relief of the original picture. Up—down mirrored pictures generally exposed a flattening of the depth relief, as compared with the relief of the original picture: Average slope values measured 0.218, 0.332, and 0.348 for A.D., E.C., and H.R., respectively (Table 1). The flattening of

the up—down mirrored pictorial relief over all poses is also clearly noticeable when the profile views of the reconstructed pictorial reliefs are looked at. Left—right mirrored pictures showed no significant depth scalings to their originals (average slope values of 1.089, 0.930, and 0.887 for A.D., E.C., and H.R., respectively; Table 1; Figure 4).

Affine regression. However, as can be noticed in the figure of the pictorial reliefs (Figures 4A–4C), depth scaling (depth compression or depth dilation) cannot cover all the differences occurring between conditions.

Table 1
Raw Regression Weights of the Depths (Straight Regression)

		Pose					
Participant	Mirror-Reflection	060	090	120	220	250	280
A.D.	LR	1.270***	1.089***	0.887***	1.081***	1.157***	1.048***
	UD	-0.142	-0.205	0.226**	0.561***	0.460***	0.410***
E.C.	LR	0.796***	0.789***	0.873***	0.922***	1.112***	1.088***
	UD	0.825***	0.313**	0.855***	0.301***	-0.165**	-0.139
H.R.	LR	0.804***	0.967***	1.171***	0.860***	0.694***	0.825***
	UD	0.385***	0.261***	0.415***	0.269***	0.254***	0.502***

Note—LR, left-right mirror reflection; UD, up-down mirror reflection. *p < .05. **p < .01. **p < .0001

When the pictorial reliefs from the original and the up-down mirrored pictures in Figure 4 are compared, they seem to undergo more complex global slant changes. A regression of z (OR) on z (mirrored) over all locations, indicating to what extent there is a depth offset and a depth scaling between conditions, considers only the depth dimension. Straightforward regression analyses are therefore unable to *completely* account for the global slant differences observed between conditions. More complex global slant differences, however, can be determined quantitatively by conducting regression analyses that involve not only the depth dimension (z-dimension), but also the image plane (x- and ydimensions). Multiple regression analyses were thus executed that included the depth values of the original picture [z](OR), as well as their pixel image coordinates (x, y), as predictor variables to predict the criterium– that is, the depth values of the mirrored pictures³ [z (LR) or z (UD)]. (The image coordinates refer to the positions on the image plane on which the depth values were calculated.) This way, a depth scaling, as well as a shear, could be revealed. (If one is not familiar with a shear, imagine, as an example, a rectangle transforming into a

parallelogram.) Explained in other, more geometrical terms, the depth values of the original picture were affinely transformed in order to approximate the depth values of the mirrored picture. This affine transformation preserves the image plane. This is a necessary condition, since the image plane has not been changed in any way. Furthermore, general affine transformations conserve planarity.

In Figure 6, scatterplots relating the affinely corrected depth values of the original pictures with the raw depth values of the mirrored pictures are presented. As compared with the raw scatterplots displayed previously (Figure 5), the affinely corrected scatterplots show strong linear trends. The coefficients of determination ranged from 89.0% to 98.4%, from 85.6% to 97.2%, and from 78.5% to 97.1% for, respectively, Participant A.D., Participant E.C., and Participant H.R. All R^2 s were found to be highly significant (for all the participants, p < .0001).

In all cases, a significant gain (p < .01) in proportion of explained variance was obtained using the affine regression model for all the participants (Figure 7). Especially, the OR–UD comparison of the pictorial reliefs

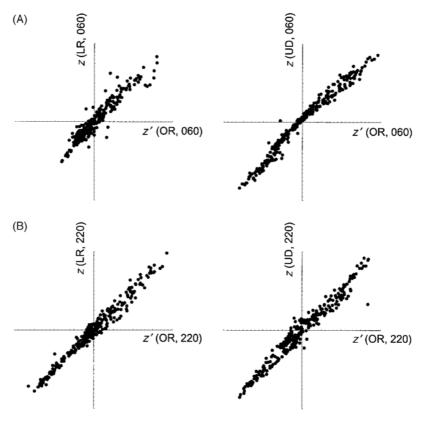


Figure 6. Scatterplots of the affinely transformed depth values of the original picture (x-axis) versus the raw depth values obtained from the left-right or the up-down mirrored picture (y-axis) for (A) pose of 60° and (B) pose of 220°, Participant E.C. z, depth values; z, affinely transformed depth values; z and z are measured in pixels. OR, original picture; LR, left-right mirrored picture; UD, up-down mirrored picture. The x- and the y-axes range from -100 to 100 pixels.

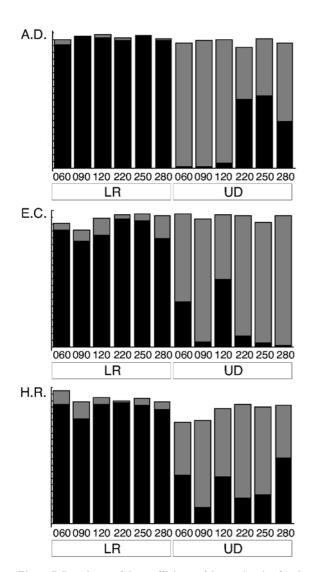


Figure 7. Bar charts of the coefficients of determination for the straight (black) and the affine (gray) regression for all poses per participant. 060, 090, 120, 220, 250, and 280 indicate the various poses in degrees. LR, comparison of original picture and left-right mirrored picture; UD, comparison of original picture and up-down mirrored picture. A.D., E.C., and H.R. refer to the participants.

showed dramatic increases in predictive power of the affine regression model, as compared with the straight model.

The gains in predictive power when moving from straight to affine regression were found to be of the same order of magnitude for all 3 participants. Similarity in magnitude of gain between the two models does, however, on no account mean that the actual affine transformations are the same for all the participants. The shears between the pictorial reliefs of the original and the mirrored pictures obtained for the different participants can be compared by depicting them in a polar coordinate representation (Figure 8). In this figure, the global shear de-

termined by the weights of the image coordinates in the regression model (see also Koenderink, van Doorn, Kappers, & Todd, 2001) is presented for each pose and for each participant. The shears for OR–LR are generally very small. Especially, the comparison OR–UD exposed significant shears. As one can notice for OR–UD, the same trends seem to occur for all 3 participants: The direction of the shear is highly similar, whereas slight differences appear for the magnitude of the shear.

Piecewise Differences

The raw scatterplots of OR–UD showed datapoints that seemed to cluster perceptually rather remarkably into a few distinct branches (Figure 5). The slope of the clusters differed: Some clusters exhibited slopes lower than one; others exposed slopes which were quite similar to one or higher. When the locations of the depth values on the image plane were examined, the observed clusters were found to correspond with specific parts of the stimulus (Figure 9).

However, the clusters observed in the raw scatterplots (depth values of original vs. depth values of mirrored picture) vanished nearly completely after affinely correcting the depth values of the original picture. For instance, the affinely corrected scatterplots presented in Figure 6 show a very strong linear relationship between the affinely transformed depth values of the original picture and the raw depth values of the mirrored version: No distinct clusters can be differentiated. When the residuals obtained after affinely transforming the depth values of the original picture were related to the image plane of the picture, it was found that the residuals were matched with different surface regions on the torso (Figure 10).

Comparing Participants

Up till now, we have dealt with differences between the conditions without extensively referring to possible interindividual differences. When qualitatively exploring the pictorial reliefs visually (Figure 4), we notice that the same trends seem to occur between conditions for the different participants. However, some differences between the pictorial reliefs of the participants seem to be the case as well. At first sight, the participants appear to have had their own "signature." In what follows, we will study the interindividual differences between the pictorial reliefs obtained from the different participants in somewhat more detail.

Global Differences

Straight regression. In order to investigate these interindividual differences more thoroughly, we first compared the depth maps of the original, left–right mirrored, and up–down mirrored pictures between participants for each pose. In general, straight regression analyses revealed high proportions of explained variance. The average R^2 for the A.D.–E.C. comparison measured .658. The A.D.–H.R. comparison had an average R^2 of .700. The

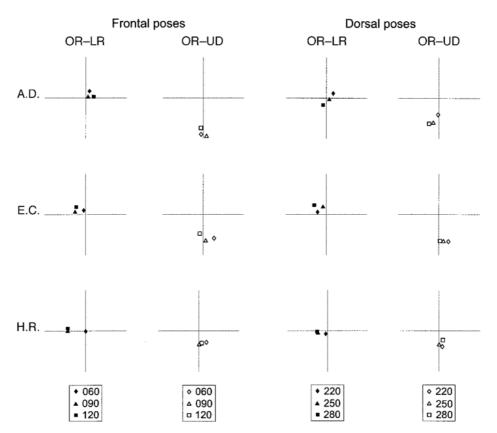


Figure 8. Polar representation of the shears between the pictorial reliefs of the original pictures and the mirrored counterparts for each participant (A.D., E.C., and H.R.). The shears represented were calculated on the basis of the weights b and c of the image coordinates x and y in the regression model (see note 3). The direction of the shear is obtained by calculating the arctangens of the ratio of the weights c and b of the image coordinates x and y. A measure of the magnitude of the shear was set between 0 and 1 by computing $\sin\{\arctan[\sqrt{(b^2+c^2)}]\}$. The x-axis and y-axis represent the directions in the image plane. Both axes range from -0.3 to 0.3, thereby presenting only a portion of the total range (-1 to 1). 060, 090, 120, 220, 250, and 280 refer to the different poses in degrees. OR-LR (OR-UD) refers to the comparison between the pictorial reliefs obtained from the original picture and the left-right mirrored (up-down mirrored) picture.

E.C.-H.R. comparison obtained an average R^2 of .744. This indicates that the pictorial reliefs of the different participants were highly similar. Some pose–condition combinations, however, showed a less pronounced linear relationship between the depth values (e.g., for A.D.-E.C., for 250UD, $R^2 = .053$; for 280UD, $R^2 = .094$; Figure 11).

Affine regression. Previous research has shown that differences concerning depth maps between participants could be ruled out by affine transforming the depths of one participant into the depths of the other participant (Koenderink et al., 2001). We were curious whether affine regression would here, too, have more predictive power than just straight regression. To this end, we chose the depth values of one participant as reference and transformed his/her data affinely toward the depth values of the other participant. Carrying out these multiple regression analyses, taking into account the depth values as well as the *x* and *y* image coordinates, a significant gain

in proportion of explained variance was observed for all pose–condition combinations [all ps < .01, except for A.D.–E.C., for 250LR, F(2,291) = 4.506; for E.C.–H.R., for 250LR, F(2,291) = 3.811; Figure 11]. (Again, the image coordinates refer to the positions on the image plane where the depth values were calculated on the basis of the participant's responses.)

Piecewise Differences

Examining the raw scatterplots of the depth values from one participant versus the depth values of the other participant, we noticed, analogous to the raw scatterplots relating the depth values obtained in different conditions, clusters of depth values that corresponded with different parts on the torso (Figure 12). Here again, correcting the depth values of one of the participants affinely toward the depth values of another participant made the clusters nearly disappear.

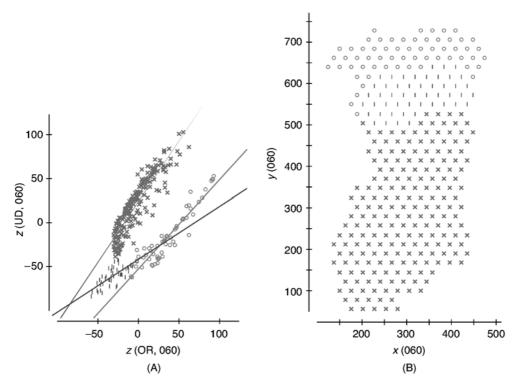


Figure 9. Piecewise differences (Pose of 60° , Participant E.C.). (A) The raw scatterplot of the depth values obtained from the original picture versus the depth values obtained from the up-down mirrored picture. Clusters are indicated by different symbols. The regression line of each cluster is presented. (B) The x,y image plane of the torso. The clusters of depth values shown in panel A are indicated on the image plane (panel B) by the same symbols as those in panel A. Note that the toxters correspond with specific surface regions on the torso. z, depth values; x,y, image coordinates; x,y, and z are measured in pixel values. OR, original picture; UD, up-down mirrored picture. Note that the origin of the axes is displaced in comparison with that of the scatterplots in Figures 5 and 6.

DISCUSSION

As was mentioned above, losing the distance-range is the inevitable result of depicting a 3-D object or scene. Infinitely many different scenes might have generated one and the same picture. Even if one has seen the real physical object/scene, one can never indicate for sure which one of the causally possible scenes is the "true one" on the basis of its 2-D depiction (Koenderink, van Doorn, Arend, & Hecht, 2002). A 2-D picture does not fully determine the actual layout of the depicted 3-D scene. Furthermore, it is impossible to have other views of the object than the view taken by the camera. Nevertheless, looking at an object or scene represented on a picture, one is still able to make a 3-D interpretation of it. In making a 3-D interpretation of a depicted object, one always has to deal with the ambiguity inherent in a 2-D depiction of a 3-D scene.

Previous studies have shown that pictorial reliefs of objects reveal differences depending on, for instance, lighting conditions (Koenderink et al., 1996a, 1996b; Todd et al., 1996), rendering conditions (Koenderink et al., 1996b), or viewing conditions (Koenderink et al.,

1994). These studies focused, in fact, on cue variations. It is not *that* surprising to have found differences in the percept, since the depth cues presented to (picture variations) or picked up by (viewing variations) the observers differed depending on the condition.

We were interested to find out whether changing the picture in such a way that its pictorial depth cues would not be affected at all would influence the pictorial relief of the depicted object. More specifically, we studied the effect of mirror-reflecting the picture on the perception of the depicted object. When mirror-reflecting the picture, the range of plausible scenes possibly corresponding with the depicted scene has not changed at all with respect to the original depiction. It is still the same infinite number. We were curious to find out whether the inherent ambiguities would be solved in the same way or differently. And if the pictorial reliefs would show differences, would they be equivalent when possible different solutions to the inherent ambiguities were taken into account?

To this end, a 3-D object (a female torso) was photographed from six different viewpoints. Subsequently, all the pictures were mirrored around the horizontal and the

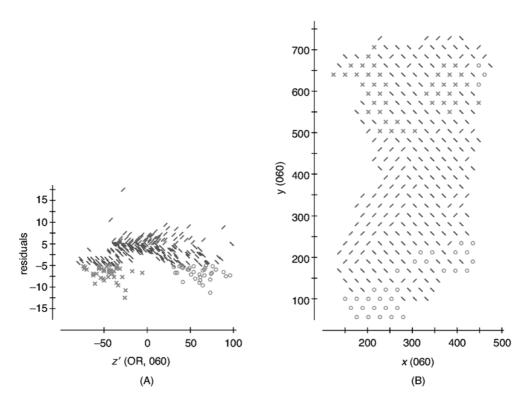


Figure 10. (A) Plot of the depth values of the original picture affinely transformed toward the depth values of the up-down mirror-reflected picture (Pose of 60° , Participant E.C.) versus the residuals. A middle band of residuals between -5 and 5 was indicated. Consequently, the remaining distinct clusters of residuals were marked by different symbols. (B) The x, y image plane of the torso. The clusters of residuals in panel A are indicated on the image plane (panel B) by the same symbols as those in panel A. Note that the clusters correspond with specific surface regions on the torso. z, affinely transformed depth values. The depths were measured in pixel values. OR, original picture. Note also that the origin of the axes is displaced in comparison with that of the scatterplots in Figures 5 and 6.

vertical axes. The participants were presented each picture, one at a time and in random order on a computer screen. Their task was to perform local attitude settings by adjusting a thumbtacklike figure so that it seemed to be painted onto the object's surface. In this way, we externalized the 3-D interpretation of the depicted object. The pictorial reliefs of the torso were constructed for all the pictures. A pairwise comparison of the pictorial reliefs of the original picture and the left—right or up—down mirrored version was conducted. We found that the pictorial reliefs derived from the raw data were dissimilar. Especially, the pictorial reliefs of the up—down mirrored pictures were strikingly different, as compared with their originals.

Depth Scaling

A classical relief ambiguity already described by Adolf Hildebrand (1918) is depth scaling. A picture of an object does not fully specify the depth ratios in the object. Depth scaling manifests itself as a depth stretch—that is, depth dilations or contractions.

First, we examined to what extent the differences between the pictorial reliefs of the original picture and the mirrored counterpart could be explained solely by this kind of depth scaling. When the depth maps calculated from the raw data were considered, flattened depth reliefs were generally obtained when the participants were looking at the up—down mirrored pictures of the torso, in comparison with the original pictures. This finding is in accordance with many art theoretici (e.g., Edwards, 1979), advising that artists should look upside down at the object of interest in order to have a better judgment of its compositional characteristics.

Furthermore, comparison of the pictorial reliefs revealed depth scaling not only in a global sense, but also in a piecewise manner. The specific depth-scaling differences corresponded with parts on the object. In the present study, the difference in orientation of these parts of the object apparently played a role. A similar finding is described in Todd et al. (1996): Changing lighting conditions showed piecewise depth-scaling effects on the pictorial relief. The various illumination conditions generated differences in shadowing and cast shadows, which were possibly of importance in explaining the differences in the perception of the surface relief.

Comparison of the pictorial reliefs extracted from the original pictures with their left—right mirrored versions did not reveal significant depth-scaling differences. This

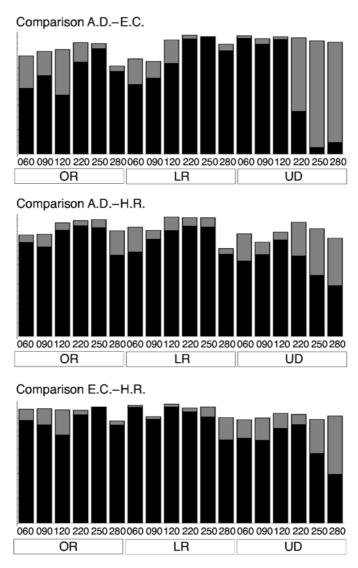


Figure 11. Bar charts of the coefficients of determination for the straight (black) and the affine (gray) regression of all possible combinations of participants for all poses and conditions. 060, 090, 120, 220, 250, and 280 indicate the various poses in degrees. OR, original picture; LR, left-right mirrored picture; UD, up-down mirrored picture. A.D., E.C., and H.R. refer to the participants.

finding is not really surprising, since we may assume that the torso is a vertically oriented object. Therefore, mirror-reflecting the picture around the vertical axis—and thus keeping the top and the bottom unchanged—possibly has less influence on the percept of the depicted object than does up—down reflection (reflection around the horizontal axis).

However, the ambiguities inherent in a picture include more than the classical depth-scaling aspects. Pure depth scaling of objects/scenes are not the only transformations causally producing one and the same photograph. Scenes "rotated" by the laws of pictorial space result in the same picture as well (Belhumeur, Kriegman, & Yuille, 1999; Koenderink & van Doorn, 2003; Koenderink, van Doorn, Kappers, & Todd, 2000; Koenderink et al., 2001; Langer & Bülthoff, 2000). As can be shown theoretically, the latter ambiguity is an extension of the group of ambiguities defined by Hildebrand (Koenderink & van Doorn, 2003).

Shear and Depth Scaling

In the present study, we investigated whether this extended group of affine ambiguity transformations could explain the differences between the pictorial reliefs. Taking the pixel coordinates of the frontoparallel plane into account, as well as the depth values obtained from the

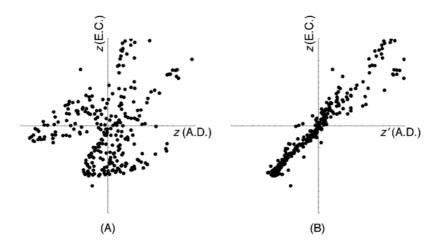


Figure 12. (A) A raw scatterplot of the depth values of Participant A.D. versus the depth values of Participant E.C. (Pose of 280°, up—down mirrored picture). (B) The depth values of Participant A.D. were affinely corrected toward the depth values of Participant E.C. (Pose of 280°, up—down mirrored picture). Note that after affine correction, the clusters observed in the raw scatterplot in panel A have nearly completely disappeared. z = depths; z° , affinely transformed depth values; the depths were measured in pixel values.

original picture, both a shear component and a depth scaling were considered. These predictor variables did amazingly well at predicting the depth values obtained from the up-down mirrored picture. Indeed, the latter model generally performed significantly better than the model accounting for depth scaling only, strongly suggesting that a shear took a place between the pictorial reliefs of the original picture and its mirrored counterpart. A similar finding was observed when the left-right mirrorreflected pictures were examined, although in this case the differences between the two models tended to be much less. Thus, for both reflection conditions, after applying an affine ambiguity transform including a shear as well as a depth scaling, the depth maps of the original and the mirrored pictures became nearly identical. The ambiguity relationship was thus certainly not restricted to a pure depth scaling.

In previous research, the pictorial reliefs obtained from three different tasks probing the percept were compared (Koenderink & van Doorn, 2003; Koenderink et al., 2000, 2001). Although no variation whatsoever concerning the stimulus material was conducted, the ambiguities incorporated in the picture seemed to be solved differently for each task. The results for the different tasks were affinely related, in a way similar to that in the present study.

A shear found when moving from one (say, from the original picture) to another (say, to the mirrored picture) pictorial relief can be understood as a global attitude change. The pictorial relief obtained from one condition seems to be considerably changed in orientation, as compared with the other pictorial relief. The only plausible interpretation is that when presented a mirrored picture, the observer changed his viewpoint so that he was look-

ing at the local surface in a different way than when he was looking at the original. However, the physical viewpoint of the observer remained unchanged during the experiment. Therefore, the observer seems to *mentally* have relocated his viewpoint (for similar mental viewpoint interpretations, see Koenderink & van Doorn, 2003; Koenderink et al., 2000, 2001).

CONCLUSION

Although we did not change anything about the inherent ambiguities in the picture by mirror-reflecting the picture, the pictorial reliefs seemed to be quite different for original versus mirrored versions. Shape perception of depicted objects does, therefore, not seem to be symmetrical. However, the differences between the pictorial reliefs of the original and the mirrored versions could be diminished drastically by conducting an affine transformation correction, taking into account not only the depth values, but also the picture plane coordinates. After being corrected in this way, the pictorial reliefs were nearly the same. The differences in the pictorial reliefs could thus be completely explained by different solutions to the inherent picture ambiguities.

REFERENCES

Belhumeur, P. N., Kriegman, D. J., & Yuille, A. L. (1999). The basrelief ambiguity. *International Journal of Computer Vision*, **35**, 33-44. Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

EDWARDS, B. (1979). Drawing on the right side of the brain: A course in enhancing creativity and artistic confidence. Los Angeles: Tarcher.GIBSON, J. J. (1979). The ecological approach to visual perception.Boston: Houghton Mifflin.

HILDEBRAND, A. (1918). Das Problem der Form in der bildenden Kunst

- (3rd ed.) [The problem of form in painting and sculpture]. Strassburg: Heitz
- KOENDERINK, J. J., KAPPERS, A. M. L., POLLICK, F. E., & KAWATO, M. (1997). Correspondence in pictorial space. *Perception & Psychophysics*, 59, 813-827.
- KOENDERINK, J. J., VAN DOORN, A. J. (2003). Pictorial space. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), Looking into pictures: An interdisciplinary approach to pictorial space (pp. 239-403). Cambridge, MA: MIT Press.
- KOENDERINK, J. J., VAN DOORN, A. J., AREND, L., & HECHT, H. (2002). Ecological optics and the creative eye. In D. Heyer & R. Mausfeld (Eds.), Perception and the physical world: Psychological and philosophical issues in perception (pp. 271-304). New York: Wiley.
- KOENDERINK, J. J., VAN DOORN, A. J., CHRISTOU, C., & LAPPIN, J. S. (1996a). Perturbation study of shading in pictures. *Perception*, 25, 1009-1026.
- KOENDERINK, J. J., VAN DOORN, A. J., CHRISTOU, C., & LAPPIN, J. S. (1996b). Shape constancy in pictorial relief. *Perception*, 25, 155-164.
- KOENDERINK, J. J., VAN DOORN, A. J., & KAPPERS, A. M. L. (1992). Surface perception in pictures. Perception & Psychophysics, 52, 487-496.
- KOENDERINK, J. J., VAN DOORN, A. J., & KAPPERS, A. M. L. (1994). On so-called paradoxical monocular stereoscopy. *Perception*, 23, 583-594
- KOENDERINK, J. J., VAN DOORN, A. J., & KAPPERS, A. M. L. (1995). Depth relief. *Perception*, 24, 115-126.
- KOENDERINK, J. J., VAN DOORN, A. J., KAPPERS, A. M. L., & TODD, J. T. (2000). Directing the mental eye in pictorial perception. In B. E. Rogowitz & T. N. Pappas (Eds.), *Proceedings of SPIE: Vol. 3959. Human Vision and Electronic Imaging V* (pp. 2-13). Bellingham, WA: International Society for Optical Engineering.
- KOENDERINK, J. J., VAN DOORN, A. J., KAPPERS, A. M. L., & TODD, J. T. (2001). Ambiguity and the "mental eye" in pictorial relief. *Perception*, 30, 431-448.
- Langer, M. S., & Bülthoff, H. H. (2000). Measuring visual shape using computer graphics psychophysics. In B. Péroche & H. Rushmeier (Eds.), Rendering Techniques 2000: Proceedings of the Eurographics Workshop on Rendering Techniques (pp. 1-9). Vienna: Springer-Verlag.
- MARR, D., & NISHIHARA, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, **200**, 269-294.
- PALMER, S. E., ROSCH, E., & CHASE, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135-151). Hillsdale, NJ: Erlbaum.
- PIRENNE, M. H. (1970). Optics, painting and photography. Cambridge: Cambridge University Press.
- ROCK, I., & DrVITA, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, **19**, 280-293.

- Rock, I., DiVita, J., & Barbeito, R. (1981). The effect on form perception of change of orientation in the third dimension. *Journal of Experimental Psychology: Human Perception & Performance*, 7, 719-732.
- SHEPARD, R. N., & COOPER, L. A. (1982). Mental images and their transformations. Cambridge, MA: MIT Press.
- STEWART, I., & GOLUBITSKY, M. (1993). Fearful symmetry: Is God a geometer? London: Penguin.
- TARR, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of threedimensional objects. *Psychonomic Bulletin & Review*, 2, 55-82.
- TARR, M. J., BÜLTHOFF, H. H., ZABINSKI, M., & BLANZ, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, **8**, 282-289.
- THOULESS, R. H. (1931). Phenomenal regression to the real object. *British Journal of Psychology*, **21**, 339-359.
- TODD, J. T., KOENDERINK, J. J., VAN DOORN, A. J., & KAPPERS, A. M. L. (1996). Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces. *Journal of Experimental Psychology: Human Perception & Performance*, 22, 695-706.
- VAN DOORN, A. J. (2000). Shape perception in different settings. In B. E. Rogowitz & T. N. Pappas (Eds.), *Proceedings of SPIE: Vol. 3959. Human Vision and Electronic Imaging V* (pp. 697-708). Bellingham, WA: International Society for Optical Engineering.
- VANRIE, J., WILLEMS, B., & WAGEMANS, J. (2001). Multiple routes to object matching from different viewpoints: Mental rotation versus invariant features. *Perception*, 30, 1047-1056.
- WAGEMANS, J., VAN GOOL, L., LAMOTE, C., & FOSTER, D. H. (2000). Minimal information to determine affine shape equivalence. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 443-468.

NOTES

- 1. z(LR or UD) = a + bz(OR), with z depth values measured in pixels
- 2. 060, 090, 120, 220, 250 and 280 refer to the different poses measured in degrees with respect to a specified reference direction; OR, LR, and UD refer to the original, left-right, and up-down conditions.
- 3. z(LR or UD) = a + b x + c y + d z(OR), with x, y image coordinates and z depth values; x, y, and z are measured in pixels.
- 4. z(Participant 2) = a + b x + c y + d z(Participant 1), with x, y image coordinates and z depth values; x, y, and z are measured in pixels.

(Manuscript received October 8, 2002; accepted for publication April 8, 2003.)