Perception, 2001, volume 30, pages 431 – 448

DOI:10.1068/p3030

Ambiguity and the 'mental eye' in pictorial relief

Jan J Koenderink, Andrea J van Doorn, Astrid M L Kappers
Helmholtz Institute, University of Utrecht, Buys Ballot Laboratory, Princetonplein 5, PO Box 80 000, NL 3508 TA Utrecht, The Netherlands; e-mail: j.j.koenderink@phys.uu.nl

James T Todd
Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, OH 43210, USA

Received 5 January 2000, in revised form 28 September 2000

Abstract. Photographs of scenes do not determine scenes in the sense that infinitely many different scenes could have given rise to any given photograph. In psychophysical experiments, observers have (at least partially) to resolve these ambiguities. The ambiguities also allow them to vary their response within the space of 'veridical' responses. Such variations may well be called 'the beholder's share' since they do not depend causally on the available depth cues. We determined the pictorial relief for four observers, four stimuli, and four different tasks. In all cases we addressed issues of reliability (scatter on repeated trials) and consistency (how well the data can be explained via a smooth surface, any surface). All data were converted to depth maps which allows us to compare the relief from the different operationalisations. As expected, pictorial relief can differ greatly either between observers (same stimulus, same task) or between operationalisations (same observer, same stimulus). However, when we factor out the essential ambiguity, these differences almost completely vanish and excellent agreement over tasks and observers pertains. Thus, observers often resolve the ambiguity in idiosyncratic ways, but mutually agree — even over tasks — in so far as their responses are causally dependent on the depth cues. A change of task often induces a change in 'mental perspective'. In such cases, the observers switch the 'beholder's share', which resolves the essential ambiguity through a change in viewpoint of their 'mental eye'.

1 Introduction

'Pictorial relief' is a term that indicates the shape of surfaces in 'pictorial space', and 'pictorial space' refers to the three-dimensional spatial impression obtained when one looks at two-dimensional pictures. In this paper, the pictures are monochrome photographs that depict pieces of sculpture in an obvious manner. They are straight illustrations rather than artistic renderings. The observer looks 'into' pictorial space instead of (or, rather, at the same time as) at the picture surface. Pictorial surfaces appear as (usually curved) boundaries of opaque objects.

Although one often identifies the pictorial surface with the surface of an actual object depicted, such an identification is strictly vacuous. This is immediately clear in the case of renderings of nonexisting objects — paintings of unicorns, say. The same holds true in the case of photographs though. Whereas it is true that at the time of the exposure the scene in front of the camera was actually there, this in no way guarantees that the photograph specifies that scene. In fact, an infinity of physical scenes might conceivably have produced the same photograph. This is especially evident in the case of cinematography where, routinely, scenes as experienced by the observer never existed in reality. For instance, a 'palace' is not necessarily anything beyond a cardboard facade.

For a number of depth cues, the nature of the relief ambiguity is formally well understood (Belhumeur et al 1997; Koenderink and van Doorn 1997). The ubiquity of such ambiguities implies that pictorial relief is only partially causally dependent on the fiducial scene.

In the final analysis, pictorial relief can only be defined *operationally.* Moreover, in view of the ambiguities, one should not be surprised when different psychophysical operationalisations yield different results. From the perspective of vision science, this is to be considered an advantage rather than a drawback, because the variation of pictorial relief with probing methods yields a powerful handle on the mechanisms underlying pictorial space perception.

Pictorial relief necessarily involves a stimulus, an observer, and a task. The task is the instrument used to operationalise the relief. Another important feature of the task is that it forces a unique response. The observer *has* to resolve the ambiguity left by the image structure, whereas this does not necessarily happen in everyday perceptions. Thus, the result depends on all three factors: the stimulus, the observer, and the task. So far, the focus of most work has been on the properties of the stimulus, that is on the type of scene, the rendering, and the viewing conditions. These factors are often summarised in the form of the major 'depth cues' involved. A few operationalisations have also been studied, but rarely have results obtained with different methods been compared quantitatively. That the observer is a key factor is clear; one often speaks of the 'beholder's share' in visual experience. Yet it is rare to see the beholder's share quantified, except in very reduced situations. In psychophysical experiments the beholder's share is generally considered a flaw and one tries to minimise it. However, because most stimuli only specify the scene ambiguously, this is not necessarily the best approach.

When two methods differ in results, then at least one of them is generally considered to be 'flawed'. This reflects the mistaken notion that every 'good' method should reflect 'reality'. Major aspects of the method are whether it involves local or global structure, and whether the subject has to judge depth (zeroth order), surface orientation (first order), or perhaps curvature (second order). These define a lower bound on the extent of the ambiguities. Thus, a change of method typically induces a change in the depth cues selected to solve the task, and thus induces a different causal dependence of the response on the optical structure. One generally hopes that most observers are somehow equivalent, that is to say, they yield similar results at least to the extent that the responses depend causally on the image structure. To the extent that this is not the case, the cause for variability has to be found in the observers themselves and one should not be amazed when this residual variability is highly idiosyncratic and cannot be easily predicted.

There is an obvious need for extensive studies covering a broad spectrum of *stimuli* and *tasks*, and over a large number of *observers*. Such studies are slow to accumulate owing to both practical and conceptual difficulties. Notable contributions were made by Glennerster et al (1996), Ernst et al (2000), and Haffenden and Goodale (2000). In this study the aim has been set rather modestly, yet we know of no more extensive study. We compare a number of (four) photographs of abstract sculpture over a number of (again four) different tasks. Four observers participated in the study. In order to enable fair comparison, the stimuli were not too different. They all depict generic egg shapes with various appendages and minor surface undulations. A major variation is to be found in the tasks. These differ with respect to the *order*, that is to say depth or surface orientation, and *scope*, that is to say local sampling or more global assessment of shape. These variations reflect our current interest in the nature of the mechanisms underlying pictorial relief. Some pertinent questions are:

- Is relief primarily to be thought of as a depth map or as an array of surface orientations?
- Is the very concept of a map misplaced to begin with, and should one think in terms of more global (not necessarily topographically organised) entities?
- How do observers handle ambiguity? Typically the task will force them to resolve the ambiguity, but how is up to them.

2 Methods

2.1 *Observers*

The observers were the authors AD, AK, JK, and JT. The observers were naïve with respect to the stimuli in the sense that they were only familiar with the depicted objects of art via the actual photographs used in the experiment. The observers ranged in age from middle aged to early fifties. Observers AD and AK are female, observers JK and JT are male. Observers AD and JT are slightly myopic, observer JK is presbyopic, only observer AK is emmetropic. Observers used their own correction when required. All had (corrected) normal visual acuity and no known visual abnormalities.

2.2 *Stimuli*

The stimuli (see figure 1) were monochrome photographs scanned from an art book, converted to high-quality photographic renderings presented on a computer screen. Since we did not take the photographs ourselves it is unknown how the final grey scale would map on the actual range of radiances in the scene. This is, of course, typically the case when one views pictures in books, on television, or in the cinema. It is irrelevant to our present purpose. All objects were photographs of pieces of sculpture by Constantin Brancusi (Geist 1975). They were selected because the objects are of the ‘abstract’ variety and are all roughly generic egg shapes with possible appendages and relatively minor surface modulations. On the whole, all objects are rather smooth. Locally, the objects are all similar, namely smooth, mostly convex surface patches. Globally, they differ in various ways. Some of the objects are slightly flattened and photographed in different orientations—some more frontally, others more obliquely. Data on focal length and so forth are not available. Pictures were displayed upon a 21-inch monitor and were of full height. Viewing distance was 75 cm. Head position was fixed with a chin rest.

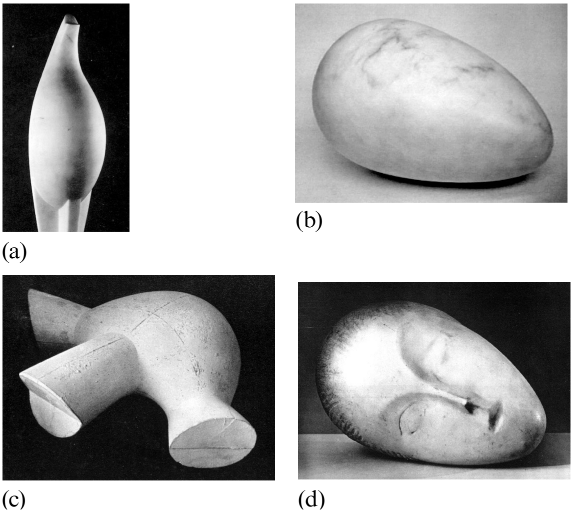

Figure 1. Stimuli used in the experiment. These are photographs of sculptures by Constantin Brancusi. (a) “Maiastra” (1915 – c. 1930), Geist (1975) catalogue 100; (b) “Sculpture for the blind” (1916), Geist (1975) catalogue 108; (c) “The turtle” (c. 1943), Geist (1975) catalogue 229; (d) “Sleeping muse” (1909 – 10), Geist (1975) catalogue 71.

Stimulus (a), “Maiastra” (1915 – c. 1930), is an abstract rendering of a bird. It is now at the Philadelphia Museum of Art. The material is white marble. The shape is elongated, with a roughly ellipsoidal (thus elliptical) ‘belly’, and a hyperbolic ‘neck’. Notice that an illuminance minimum marks the position of the parabolic curve that separates the elliptical and hyperbolical regions. The pose is ‘frontal’, ie the view is at right angles from the major axis.

Stimulus (b), “Sculpture for the blind” (1916), is an abstract egg shape. It is now at the Philadelphia Museum of Art. The material is again white marble. Slight volume texture is more evident than for stimulus (a). The object is convex (elliptical) throughout. The view is fairly 'frontal', though a slight obliquity is evident.

Stimulus (c), "The turtle" (c. 1943), is the abstract rendering of a walking turtle. It is at the Musée National d'Art Moderne, Paris: Brancusi studio. The material is plaster, much duller than the marble surfaces. Thus, the effects of shading are quite distinct from those on the marbles. Some 'seams' of the mould are evident. (The original was carved from wood.) These are important depth cues and they emphasise the plane of bilateral symmetry. This object has a quite different aspect ratio from the others in that it is pronouncedly flattish. It clearly has bilateral symmetry. The pose is very oblique—the object evidently lies on a horizontal plane that slopes away from the viewer.

Stimulus (d), "Sleeping muse" (1909 – 1910), is the rendering of an abstract head. It is at the Hirshhorn Museum and Sculpture Garden, Smitsonian Institution at Washington, DC. The material is again marble, but with a quite different finish from the other items. Though the shape is very similar to that of stimulus (b), the modulations of the facial features make it rather more complex. There are pronounced shadows of eyebrows, nose, and lips. The view is a frontal one.

For all stimuli we found the outer perimeter in the image and triangulated the interior with a regular array of vertices (see figure 2) such that the number of faces of the triangulations was about 150 triangles. The actual number of vertices was between 59 and 68, that of the edges was between 82 and 100, and that of the faces was between 140 and 167. The edge length (in pixels) was between 29 and 52. Collinear ranges of vertices ran in the horizontal (0°), 60°, and 120° directions; for each direction there were between 8 and 16 such collinear ranges containing from 1 to 10 vertices.

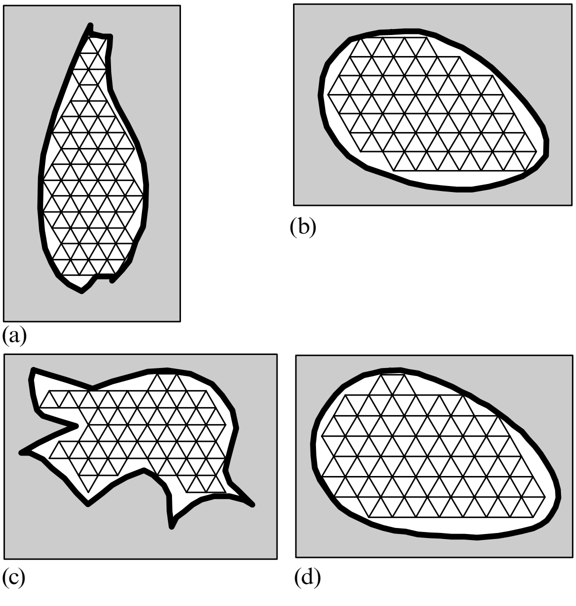

Figure 2. Triangulations of the stimuli: The thick closed curve is the contour of the object. All vertices of the triangulation are inside its interior. The edges of the triangulation are all of equal length and have orientations of 0° (horizontal), 60°, and 120°. Thus, all faces of the triangulation are congruent equilateral triangles, either 'base up' or 'base down'. The triangulation is connected in all cases.

2.3 *Tasks*

We deployed four different tasks, namely: 'gauge-figure adjustment', denoted 'AT' for 'attitude probe'; 'pairwise depth difference comparison', denoted 'PC' for 'pairwise comparison'; 'cross-section reproduction with horizontal replica', denoted 'CH' for 'cross-section horizontal'; and 'cross-section reproduction with parallel replica', denoted 'CP' for 'cross-section parallel'. With these tasks we probed different aspects of pictorial shape. Here is a detailed description of these tasks.

Gauge-figure adjustment is a task we have described in detail before (Koenderink et al 1992). We have previously used this task to study the effects of changing viewing conditions (Todd et al 1996).

This is a local-surface-orientation probing task. We superimpose a 'gauge figure' in red wireframe rendering over the monochrome (grey tones) photograph. The gauge figure is centred at the point where the sample is to be taken. This location is automatically set by the program and is taken from a randomised list of a few dozen locations. The observer has no control over position. The gauge figure is the projection of a circle with a line segment drawn from the centre of the circle at right angles to its plane. The gauge figure looks much like a thumbtack (drawing pin). The orientation and form (parameterised by the slant and tilt angles) of the gauge figure are controlled by the subject. The task is to adjust the gauge figure such that the circle appears as if painted on the pictorial surface with the line segment sticking outwards. When a natural user interface is provided, the observer can perform the task in a few seconds and typically feels quite confident about the setting. We collected three settings on each of a few dozen vertices in the course of a single session.

The *pairwise depth-comparison* task is also quite local, but it differs from the gauge-figure task in that a depth-comparison judgment is required, not a surface-orientation judgment. We have used this task before and have compared it with the gauge-figure-setting task (Koenderink et al 1996).

This is a local-depth-comparison task. The points whose depths are to be compared are marked by a pair of coloured dots superimposed over the monochrome photograph. The locations of the dots are chosen from the same collection of locations at which the gauge figure might appear. Globally, the locations are vertices from a regular hexagonal array. In this task, we always select two locations that are neighbours on the hexagonal grid, and thus are quite close together. This is why we denote the task 'local' although actually two slightly different locations are to be compared. Pairs of points of slightly different colour appear in random order. The observer has no control over the presentation. The observer is required to indicate which of the two is closer. In order to cancel out the possible effect of dot colour on perceived depth we always averaged over mutually reversed pairs.

The *cross-section reproduction with horizontal replica* task (see figure 3) is a task with which we have no prior experience whatsoever. A similar method was pioneered by Frisby et al (1995). Our method is perhaps best characterised as a perversion of theirs. We asked the observers to indicate the shape of certain cross sections of the pictorial surface. By 'cross section' we mean the shape of the curve of intersection of the pictorial surface with a plane of infinite slant (or, if one prefers to measure slant

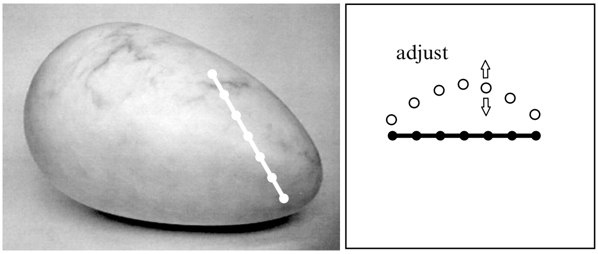

Figure 3. The method of cross-section adjustment. The subject is confronted with the stimulus (left) with a superimposed set of collinear points. The subjects adjust the corresponding points in the righthand window such as to indicate the cross section. That is to say, the shift orthogonal to the line is interpreted as depth. This is the CH task, thus the 'replica', that is the (initially straight) curve in the response window, is horizontal and the adjustment is in the vertical direction. This is the case regardless of the fact that the orientation of the section in the image window is often oblique.

as an angle, a slant of 90°). The sectioning surface thus appears in the picture as a line. This is indeed how we implemented the method. We indicated a straight line (it was simply superimposed on the picture) and required the observer to 'sketch' (to be explained later) the shape of the intersection. Because our fiducial locations were the vertices of a hexagonal grid (see above), we used lines sloped in the picture plane at 60° intervals. Such lines contain a sequence of vertices and we asked the subjects to indicate the depths of these vertices. Of course, we implemented the depth indication in such a way that the observers were primarily aware of setting the shape of the cross section, rather than depths at individual locations. Frisby et al (1995) rather elegantly used a number of rods that could be pushed away from the observer by various degrees. The endpoints of these rods then defined the cross section. We implemented a variation of this method on the monitor screen. The nominal task of the observer is to reproduce, in a separate window showing the carrier line, the pictorial profile, or cross section, of the pictorial surface along the cut indicated by a straight line superimposed over the image. The observer has to drag the points on a straight line at right angles to this line such as to reproduce the cross section. In this task the replica is always horizontal, regardless of the direction of the cross section in the image. The observer drags the points on the carrier such as to indicate the cross-sectional curve. The actual distance from the carrier was considered irrelevant; only the slope and curvature with respect to the carrier were significant and the observers were made fully aware of that.

Notice that this task is a global one, with the actual position of any single point being irrelevant, and addresses depth rather than surface orientation.

The *cross-section reproduction with parallel replica* task (see figure 4) is in most respects similar to the cross-section reproduction with horizontal replica task. The single difference is that the replica is always parallel to the direction of the cross section in the image. Although this may appear only a minor variation, all observers experience the two cross-sectional reproduction tasks as *very* different. Since the results obtained with these tasks are also quite different in most cases, we consider these two tasks as essentially distinct. Further analysis fully bears this out (see below).

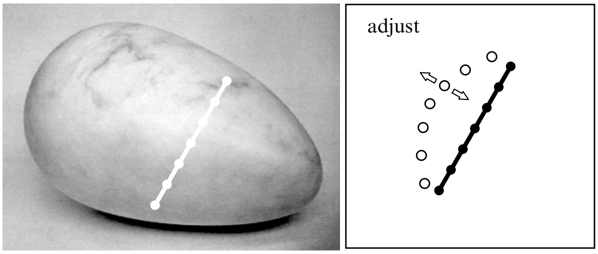

Figure 4. The method of cross-section adjustment. The subject is confronted with the stimulus (left) with a superimposed set of collinear points. The subjects adjust the corresponding points in the righthand window such as to indicate the cross section. That is to say, the shift orthogonal to the line is interpreted as depth. This is the CP task, thus the 'replica', that is the (initially straight) curve in the response window, is always parallel to the orientation of the section in the image window. The adjustment is at right angles to this orientation and thus often in an oblique direction.

2.4 *Design and data representation*

We have four stimuli [(a), (b), (c), and (d)], four tasks (AT, PC, CH, and CP), and four observers (AD, AK, JK, and JT). All observers performed all tasks repeatedly for all stimuli, with each stimulus providing many different points from which observer settings could be obtained. There were forty-eight sessions for each observer: three repetitions of each of the sixteen stimulus – task combinations. The order of the sessions for any observer was globally randomised. Per session, each single setting position was presented three times, randomised over the session. Thus we have a total of nine independent settings for each data item. Since there are dozens of data items per stimulus, we end up with thousands of lists of nine settings each. In the case of the gauge-figure adjustment, a single setting yields a pair of numbers (slant and tilt); in the case of the pairwise comparison, the result of a single case (two settings because we always have symmetrical pairs) is a pair of booleans. Thus we have a particularly rich set of data (several tens of thousands of numbers). This should amply suffice for statistical analysis to considerable depth.

Because we need a 'common currency' to make a comparison possible at all, we converted the results of all experiments to a pictorial depth map. Although this is a rational choice — indeed, there are few options — we stress the fact that using depth as the common currency does not imply that we attach a special meaning to depth as compared with surface attitude or curvature for instance. Formally such measures are fully equivalent. For instance, from the depth map we can calculate surface attitude and vice versa. We are not interested in offsets (constants of integration) here.

We particularly want to stress the point that the question of whether the brain uses depth maps as a convenient common currency between different cues is no doubt an interesting one, but it has nothing whatsoever to do with our choice.

3 Experiments and data representation

The experiments were done in sessions of roughly a quarter of an hour each. As said above, we devoted three sessions to each method and each stimulus, and the sequence was randomised overall.

In the case of the gauge-figure task the obvious first thing to do is to study the scatter in the settings. With so many samples of local surface orientation we can also construct a surface that conforms to these settings in the least-squares sense. This yields a measure of how well the data conform to a surface (any surface), we call this the 'consistency' measure. The other result is the fitted surface. This depth map is only determined up to an arbitrary depth offset. We have arbitrarily but conveniently normalised the depth maps such as to force the average depth to be zero.

The pairwise depth-comparison task was judged as considerably more difficult by all observers than gauge-figure settings since the observers often had to guess. That guesses may be necessary follows from the fact that the depth difference between a given pair of vertices might actually be zero. In the course of a single session, each pair of vertices is shown in both orders, each order ('A – B' or 'B – A') an equal number of times. Over all sessions we collect a frequency of "A is closer than B" for each ordered pair of neighbouring vertices.

Since we presented every configuration eighteen times (each pair was always presented twice, once in each order, so as to cancel out possible artifacts) we are able to study the scatter in repeated judgments. It is also possible to compute a surface that fits the data in a least-squares sense. In order to do so, one uses the psychometric curve to induce a metric. When this procedure does not yield a robust result (this would happen when the frequencies of seeing one vertex in front of another were all either zero or one hundred percent) one can at least find a global depth (partial) order. Such a procedure yields a measure of consistency—that is how well the measurements conform to *any* surface—as well as a global depth map. The consistency measure is a count of the number of 'intransitive triangles' (see below).

There are various ways to analyse the type of data yielded by the cross-section-reproduction tasks. The first thing, of course, is to compare the cross sections obtained from repeated trials. After factoring out the mean we are left with a measure of scatter in repeated settings. It is also possible to fit a surface that fits the cross sections in a least-squares sense. In such a calculation the cross sections may be moved to and fro in depth such as to fit a surface as well as possible. One obtains a measure of goodness of fit (how well the cross sections fit any surface; thus this is also a consistency measure) as well as the fitted surface or depth map. In the cross-section-reproduction tasks the measures of consistency and reliability cannot easily be separated.

Thus, the experimental results are represented as depth maps with vanishing average depth for all four tasks. Moreover, in each case we have measures of consistency (that is how well any smooth depth map represents the data) and of reliability (that is the scatter in repeated trials). These latter two measures are in task-specific formats that do not easily permit them to be compared quantitatively.

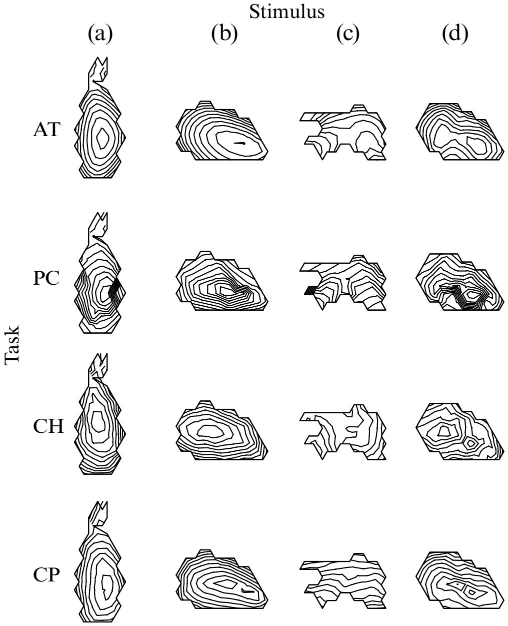

Figure 5. Pictorial reliefs for observer AD. The depth contour curves are drawn for equally spaced depth values, though the spacing is different in all cases because we drew the same number of depth contours in each diagram regardless of the total depth range. Methods AT – CP and stimuli (a) – (d) are described in the text.

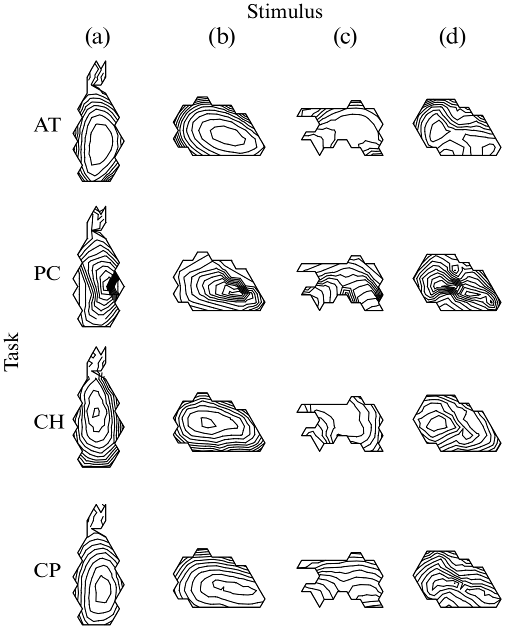

Figure 6. Pictorial reliefs for observer AK. The depth contour curves are drawn for equally spaced depth values, though the spacing is different in all cases because we drew the same number of depth contours in each diagram regardless of the total depth range. Methods AT – CP and stimuli (a) – (d) are described in the text.

In figures 5 – 8 we have collected the depth maps for all stimuli, tasks, and observers. Although only a few equal-depth curves have been drawn, these figures already yield a good global overview. It is evident that the pictorial relief may vary appreciably, especially when the task is varied. Of course, these figures show only one aspect of the data. In addition we have to consider reliability and global consistency.

Notice that the results for tasks AT, PC, and CP tend to be quite similar, whereas the result for task CH appears to be of a quite different character, especially in the case of stimulus (c). The fact that the difference is so striking between the results for the, at first glance rather similar, tasks CH and CP comes perhaps as a particular surprise (see below).

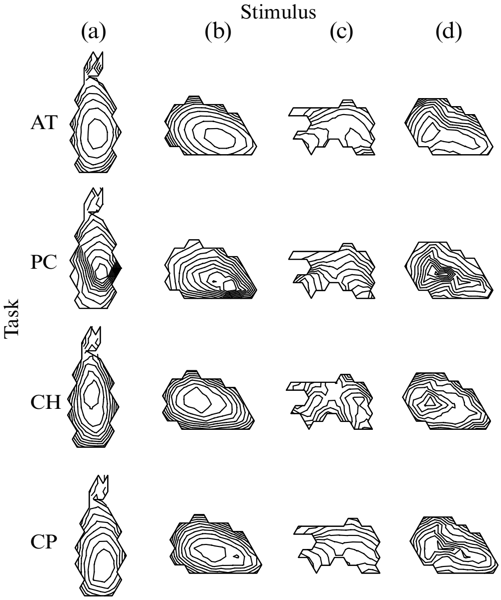

Figure 7. Pictorial reliefs for observer JK. The depth contour curves are drawn for equally spaced depth values, though the spacing is different in all cases because we drew the same number of depth contours in each diagram regardless of the total depth range. Methods AT – CP and stimuli (a) – (d) are described in the text.

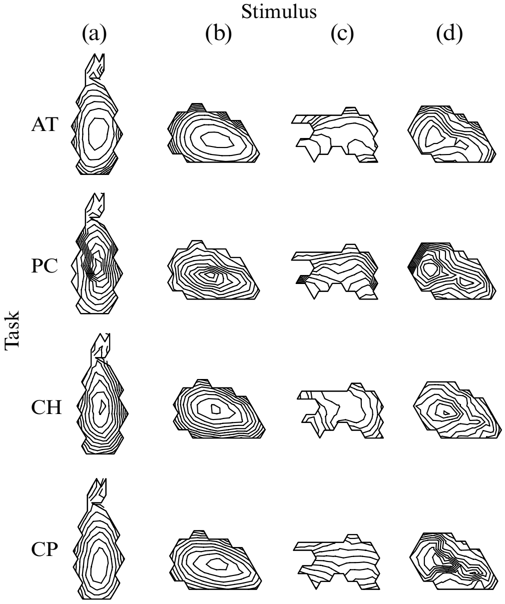

Figure 8. Pictorial reliefs for observer JT. The depth contour curves are drawn for equally spaced depth values, though the spacing is different in all cases because we drew the same number of depth contours in each diagram regardless of the total depth range. Methods AT – CP and stimuli (a) – (d) are described in the text.

3.1 *Analysis*

3.1.1 *Initial analysis.* One simple first pass at a comparison of the pictorial reliefs obtained for the various observer – task combinations is to study scatter plots of depth values for given vertices. Since the vertices belong to the triangulation of a given stimulus, such scatter plots must be made for each stimulus separately. We may plot either the depths for a pair of methods and a single observer, or the depths for a pair of observers and a single method. In either case the result may be characterised by the coefficient of determination (often denoted R^2).

Apart from the coefficients of determination we also have the linear trends, that is to say the ratios of depth ranges. We have found no clear pattern in these values. It is not the case that an appreciable part of the variance is accounted for by individual 'depth gains' for instance.

For the gauge-figure task, reliability is much like we reported before (Koenderink et al 1992). The tilts reproduce extremely well, whereas the slants show appreciable scatter. The rms spread in the magnitude of the depth gradient is 11.4%. The variance is not isotropically distributed with the variance in the gradient component in the slant direction being about 30% larger than that in the tilt direction.

The surface consistency measures are of the order of a few pixels in depth (4.5 with a spread of 2.6). This is much less than the average depth variation over a face of the triangulation, which is about an order of magnitude larger. We may conclude the results always conform closely to a coherent surface.

Because the consistency is invariably high, the depth differences over the edges of the triangulation reflect the original settings closely. A linear regression of the depth lists between observers yields high coefficients of determination for all four stimuli (see table 1). Thus, all observers produced very similar reliefs. The remaining differences apparently reflect differences between the stimuli, some are indeed qualified as 'easy' and others as more 'difficult' by the observers. The linear trend of the regression of depth lists reveals the ratios of depth range for pairs of observers. We find quite appreciable differences in depth gain (see table 2), with ratios of up to a factor of 2.2 [for stimulus (d)]. This reflects what we have found earlier: subjects are quite variable in their depth scalings though they otherwise tend to correspond closely. Such effects are perhaps to be expected since one of the ambiguity dimensions for several cues is an arbitrary depth scaling.

Table 1. Ranges of the coefficient of determination (specified as percentages) for the linear regression of depth lists between all observers for specific stimuli and tasks. In the indicated case (†), the minimum linear correlation coefficient was negative and we set the coefficient of determination to zero in order to allow fair comparison of the ranges.

Stimulus	Task			
	AT	PC	CH	CP
(a)	90 – 96	69 – 81	76 – 90	86 – 96
(b)	85 – 98	71 – 88	79 – 93	72 – 92
(c)	98	74 – 90	0† – 96	86 – 98
(d)	62 – 88	48 – 76	69 – 90	14 – 89

Results for the pairwise depth-comparison task are processed by a method we have described in detail before (Koenderink et al 1996). We may construct a depth map to account for the probability estimates (observed frequencies). A natural measure of consistency is the frequency of 'intransitive triangles', that are triangles A, B, and C such that A is closer than B, B is closer than C, and C is also closer than A.

We found that the observers repeated their judgment very accurately in repeated settings. As many as 83% to 92% of the settings were perfectly reproduced on each revisit.

Table 2. The ratios of depth ranges ('depth gains') for the linear regression of depth lists between all observers for specific stimuli and tasks. The values represent the maximum depth gain between any two observers. Notice that depth ranges are positive and necessarily in excess of unity [for a range $a < 1$ (say) can also be written $1/a > 1$ when the sequence of observers is reversed]. This applies to 'normal' cases, that is when the results are qualitatively similar, whereas in the indicated case (*) the depth ratios ranged from -20 (thus negative!) to 2.21 and so we entered a wild card in the table in order to mark that things apparently went haywire. In this case the linear regression yields a negative linear correlation coefficient (see table 1).

Stimulus	Task			
	AT	PC	CH	CP
(a)	1.39	1.67	1.89	1.47
(b)	1.43	1.41	1.41	1.49
(c)	2.13	1.10	*	3.51
(d)	2.17	1.51	2.00	4.54

This means that the psychometric curve cannot be used to full advantage in the surface reconstruction, and the computed depths are in fact little more than a depth *order*. The coefficient of determination for a linear regression of the depth lists may therefore be expected to be on the low side. However, the coefficients of determination are in fact quite appreciable (see table 1). The ratios of the depth ranges are similar to the gauge-figure case (see table 2). We find depth gains between observers of up to a factor of 1.67 [for stimulus (a)].

The number of intransitive triangles varied between 0 and 3, with an average of 0.8, that is less than 1% of the total number of triangles. Clearly the judgments conform closely to a coherent surface.

Two observers rated the cross-section-reproduction horizontal task as a reasonable task to perform, though less easy and intuitive than the gauge-figure task. The other observers rated the task very difficult and indeed unnatural. Curiously, despite these strong preferences, all observers responded in qualitatively and quantitatively similar ways in this task.

Again, two observers (the other two) rated the cross-section-reproduction parallel task as a reasonable one, though less easy and intuitive than the gauge-figure task, whereas the other observers rated the task very difficult and indeed unnatural. Once again, all observers responded in qualitatively and quantitatively similar ways in this task. These data are analysed in exactly the same way as for the cross-section-reproduction horizontal task.

For both variants of the cross-section-reproduction task we find that a linear regression of depth lists for pairs of subjects yields coefficients of determination that vary widely (see table 1). The ratios of depth ranges are rather high (see table 2). For the horizontal variant we find depth gains between observers of up to a factor of 2.0, and for the parallel variant depth gains are up to a factor of 4.5 [both for stimulus (d)]. For stimulus (c) the linear regression sometimes yields negative linear correlations, in which case a comparison of depth ranges seems to make little sense indeed.

From the results of the cross-section-reproduction tasks we can construct depth maps of the pictorial surface. The results for both the CH and the CP tasks were very similar with respect to reliability and consistency. In both cases the rms residuals for the best fitting surface vary from 5 to 15 pixels (about 7 on average) (in depth). Thus the results do conform quite well to a coherent surface for both variations of the task, though the consistency seems perhaps less than for the other tasks, in so far as one may compare such different measures at all.

3.1.2 *Detailed comparison of pictorial reliefs.* In many cases (though by no means always) we find rather high coefficients of determination when we compare observers for a given task, whereas we find generally (but by no means always) low coefficients of determination when we compare tasks for any given observer. It seems likely that the difference with a change of task is due to the fact that the observer will select a different bouquet of cues in either case, whereas the difference with a change of observer must be due either to a different selection of cues, or to the inherent ambiguity of the image (thus making room for a beholder's share), or both. In order to be able to distinguish between such possibilities we need methods that enable us to factor out the effect of the inherent ambiguity.

Most 'depth cues' are not specific as to the zeroth order, that is distance (thus 'depth' cue is really a misnomer in most familiar cases), nor in fact to the first order, that is surface attitude. For instance, the shading cue yields information relating primarily to the third order. Thus neither the (average) depth, nor the surface attitude can be particularly evident in pictorial space, at least not of a small planar element in isolation. The case is different for planes of large extent, like the ground plane (Gibson 1950), but in the present setting we may safely ignore this. This fact can be used to gain a handle on the nature of the ambiguity. In the appendix we derive an analytic form for the ambiguity and show that it is an affinity that affects the depth dimension and involves the dimension of the picture plane. Intuitively this result is obvious. The ambiguity transform should conserve the dimensions on the picture plane and planarity in pictorial space. The only such transformations are certain affinities.

With this analytic form we have gained a powerful handle on the nature of the ambiguities. Suppose we have a scatter plot of depths z_1 against z_2. Instead of a straight linear regression we perform a multiple regression (with constant factor) of z_2 against z_1, x, and y (the image coordinates). We will refer to this as an 'affine regression'. The affine regression yields the best values (in a least-squares sense) of the constants in the relation $z_2 = a + bx + cy + dz_1$, thus it has the effect of factoring out the expected ambiguity. Whether this works or not can be judged by noting the increase in the coefficient of determination when we move from mere linear to affine regression. The constants can be interpreted in terms of a depth scaling, and the orientation and magnitude of a depth shear. We show a striking example in figures 9 and 10.

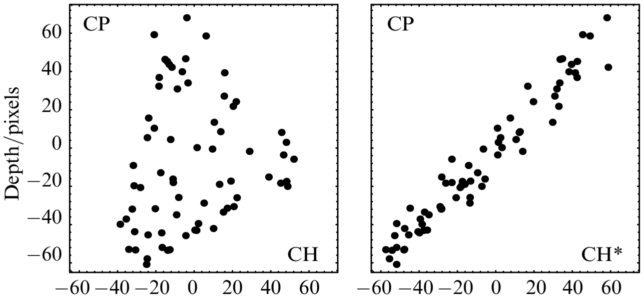

Figure 9. Scatter plots of the depth values in the case of observer AK, stimulus (a), tasks CH and CP. (i) A straight scatterplot of the depth for the two tasks. (ii) A scatterplot in which the depth values for the task CH have been affinely corrected (denoted CH*). Notice the enormous increase of coefficient of determination in these regressions. Apparently, the observer's responses are the same in both tasks, though this fact is not apparent from the straight depth values. In a case like this the straightforward linear regression of depths makes very little sense. One might think that the responses for the two tasks are hardly related. This is far from being the case though, for after discounting the effect of ambiguity, the—actually very high—correlation becomes evident.

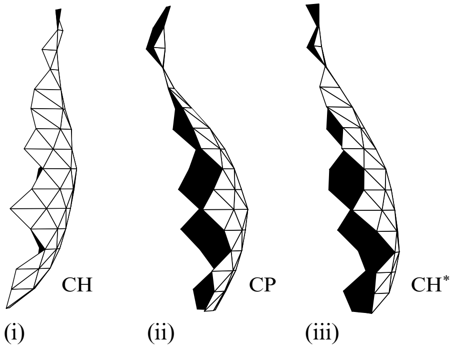

Figure 10. Renderings of pictorial relief in the case of observer AK, stimulus (a), tasks CH and CP. Vertical is the vertical in the image, left to right is the depth dimension. (i) The result for task CH, (ii) the result for task CP. Notice that these reliefs are quite different. (iii) The relief for task CH is rerendered after factoring out the ambiguity. Now the reliefs obtained in the two tasks are seen to be quite similar. This is a clear illustration of a change of 'mental viewpoint'.

In figure 11 we show coefficients of determination for the depths of pairs of methods for linear and affine regression. Of course, the values for the affine always exceed those for linear regression because, by construction, they can never be lower. Observers are grouped by row, stimuli are grouped by column. In each subgraph the sequence of bars is AT – CH, AT – CP, AT – PC, CH – CP, CH – PC, and CP – PC. In figure 12 we show coefficients of determination for the depths of pairs of observers for linear and affine regression. Methods are grouped by row, stimuli are grouped by column. In each subgraph the sequence of bars is AD – AK, AD – JK, AD – JT, AK – JK, AK – JT, and JK – JT. Clearly, the attempt to factor out the ambiguity has been extremely effective. After removal of the influence of ambiguities the coefficients of determination are invariably very high. In extreme cases, notice especially the cases for the CH – CP comparison for stimulus (c) in figure 11, the coefficient of determination goes from almost 0 to almost 1. In such cases almost all the difference between the responses was apparently due to the ambiguity.

From the results shown in figure 11 it can be concluded that the different tasks yield virtually identical results after the ambiguities have been factored out. Thus pictorial relief modulo the affine ambiguity is almost invariant against a change of task.

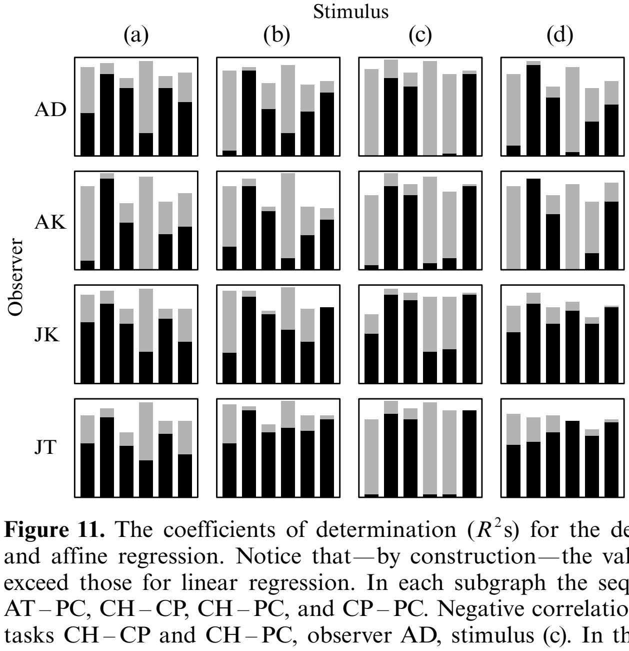

Figure 11. The coefficients of determination (R^2s) for the depths of pairs of methods for linear and affine regression. Notice that—by construction—the values for the affine regression always exceed those for linear regression. In each subgraph the sequence of bars is AT – CH, AT – CP, AT – PC, CH – CP, CH – PC, and CP – PC. Negative correlation coefficients are found for the cases tasks CH – CP and CH – PC, observer AD, stimulus (c). In these cases the coefficient of determination, of course, misrepresents the actual state of affairs. For affine regression the correlation was always positive.

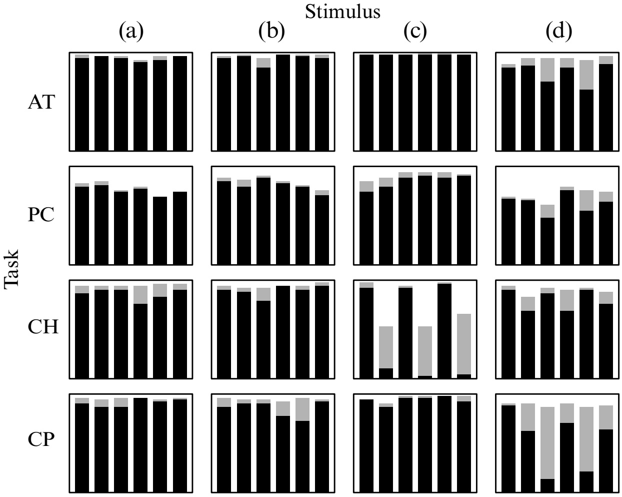

Figure 12. The coefficients of determination (R^2s) for the depths of pairs of observers for linear and affine regression. Notice that — by construction — the values for the affine regression always exceed those for linear regression. In each subgraph the sequence of bars is AD – AK, AD – JK, AD – JT, AK – JK, AK – JT, and JK – JT. Negative correlation coefficients are found for the cases task CH, stimulus (c), observers AD – JK, AK – JK, and JK – JT. In these cases the coefficients of determination, of course, misrepresent the actual state of affairs. For affine regression the correlation was always positive.

This important outcome makes the concept of 'pictorial relief' far less volatile than it appears at first glance. Ab initio one has to expect that pictorial relief, which can only be operationally defined, will depend upon the method to find it. In retrospect the actual operationalisation turns out to be nearly irrelevant. Thus 'pictorial relief'— of course, modulo the affine ambiguity — seems to have an invariant meaning that is nearly independent of the operationalisation. Notice that this is not the case for pictorial relief in the 'absolute' sense. Indeed, when you consider the CH – CP comparison for stimulus (c) you find almost completely uncorrelated 'reliefs' for all observers.

From the results in figure 12 it may be concluded that different observers often, and perhaps surprisingly so, agree in the way they resolve the ambiguity. In such cases their beholder's shares are the same. However, there are some notable exceptions. In the cases of stimulus (c), observer JK and stimulus (d), observer JT, the observers evidently resolved the ambiguity in a way that differed from the other three. All observers have the same, or very similar, pictorial relief modulo the affine ambiguity. Their responses are nearly identical in so far as they causally depend on the stimuli. The idiosyncrasy lies in the way observers resolve the ambiguity—that is to say, in their 'beholder's share'.

What is perhaps most remarkable is that the various beholder's shares are in many cases very similar. This can be seen quite well in figure 13, where the shears of all observers are plotted in the same diagram. There is one diagram for each stimulus and task comparison. Since a shear is defined through an orientation and a magnitude, it is most conveniently plotted in a polar diagram. Orientation repeats after 180°, so only half the directions need be considered. In the figure, the direction specifies the orientation of the shear, whereas the radius specifies the magnitude of the shear [on a scale from zero (no shear) to one (maximum shear); thus, all points fall in a hemi-unit circle]. The most interesting case is probably the comparison CH – CP (because these tasks appear so similar at first glance). Clearly the orientations all cluster in a rather narrow angular range. Thus, the beholder's share is a quite similar shear for

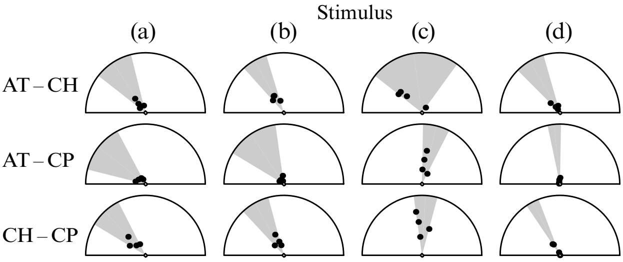

Figure 13. The depth shears for all observers and all stimuli. Both the orientation and the magnitude of the shears are plotted as points in a polar plot. Notice that the clouds of data points cluster in the orientation domain. To make this visually evident we have indicated the full orientation ranges as grey sectors.

any observer, at least as far as its direction is concerned. The extents of the shears are quite different though. These similarities have to be due to a similar interpretation of semantic image information (see section 4).

4 Conclusions

There are quite a number of important conclusions that we may draw from these experiments.

The tasks are clearly quite different in many respects. The observers rate them quite differently with respect to 'naturalness', whatever that may mean, and difficulty. By far the easiest task to perform, and also by far the task that feels most 'natural', is the gauge-figure task. In this case, the observers have to judge whether the gauge figure 'fits' the relief. The judgment is immediate and involves no overt reasoning, since the observers do not have to abstract from the perception of the image. The pairwise depth-comparison task is also easy — though boring — but feels less natural. Here the observers have to abstract from immediate perception, since the judgment 'A is closer than B' (A and B marks in the image) involves overt reasoning, be it of the simplest kind. The cross-section-reproduction tasks are not so much 'unnatural' as indirect. They involve not so much overt reasoning as the comparison of perceptions (namely the pictorial space induced by the image and the replica on the CRT screen) of quite different kinds. The tasks involve a number of 'mental transformations'. This no doubt accounts for the fact that two of the observers had a very strong preference for the parallel replica, and the other two for the horizontal replica. This apparently reflects their different ways of applying the necessary transformations.

With respect to reliability, the gauge-figure task is no doubt the most reliable, as we have reported before (Koenderink et al 1992, 1996). Since the gauge-figure task involves direct sampling of surface attitude, whereas the other tasks more directly involve depth, or at least depth differences over space, this might indicate that a representation of local surface attitudes is prior to a 'depth map'.

With respect to surface consistency, all tasks yield results that very closely approximate a coherent surface rather than a cloud of dispersed points in pictorial space. It is difficult to compare the consistency measures of these tasks quantitatively; in any case, consistency has to be judged relative to reliability. It seems to us that a reasonable conclusion is that all tasks yield pictorial reliefs that are fully consistent surfaces within the range specified by the respective reliabilities. This is one argument for the validity of the very concept of 'pictorial relief'.

Perhaps surprisingly, the pictorial reliefs are very similar indeed (coefficients of determination are generally in the 0.8 – 1.0 range) for all four tasks and for all four observers after the ambiguities have been factored out. This appears to indicate that the different observers selected the same bouquet of depth cues and arrived at their perceptions in rather similar ways. Apparently the observers may safely be said to see similar three-dimensional pictorial spaces when they look at the same photograph of a physical scene. Although this tentative conclusion is generally applied as a rather certain assumption in daily life (people examining holiday snapshots together, say), from a fundamental perspective it seems surprising that this would actually apply.

Notice that this involves very important methodological issues. First, if one uses straightforward regression in the comparison of depths in pictorial relief one is likely to be led to completely fallacious conclusions. One may conclude that responses of different observers fail to agree where they are actually very close. Second, to report local surface orientations (as in the gauge-figure task or some equivalent) in terms of surface normals is strictly senseless, because orthogonality in pictorial space is destroyed by the ambiguity transforms. To use the depth gradient seems most prudent. An ambiguity transform then implies a scaling by the depth gain and a constant offset determined by the shear of the depth gradient.

In the classical view, which is due to Hildebrand (1893/1945), the ambiguity is limited to a depth gain. This is indeed a very common phenomenon and we have reported on it earlier (Koenderink et al 1992). It appears here that this view is unduely limited though. Under certain circumstances, the shear component of the ambiguity transform can easily dominate the effects of the depth gain.

By far the larger part of the variation encountered by us in these experiments can be described in terms of depth scalings and shears. These affinities are partly idiosyncratic, as one would expect, but for the larger part appear to follow a common pattern for all observers. By and large, observers apply similar depth scalings and shears—which choice they arrive at thus seems to be determined by the stimulus. This is highly surprising (to us). It indicates that the observers apply similar priors, because the causal influences have already been accounted for by the pictorial relief modulo the affinities. Apparently these priors depend at least partly on the stimuli though.

One possible explanation might be the following. Since the images underdetermine pictorial relief, the observers may use the remaining freedom to *adjust their 'mental eye'.* This is illustrated in figure 14. They may take another perspective, subject to certain constraints. When the mental eye changes viewing direction, this may not affect the occluding boundary. For the image fully and finally specifies all that is visible; one may not look at the rear of the depicted object, say. But the change of perspective can be used to select any point as locally frontoparallel. This would of course be impossible in real space with a physical change of vantage point. In that case, again any point can be made to be locally frontoparallel, but the necessary consequence

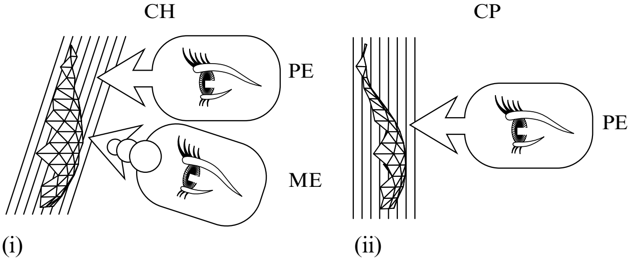

Figure 14. The shear in the case of observer AK, stimulus (a), tasks CH and CP. A view of the pictorial relief in case of (i) task CH, (ii) task CP. The viewing direction of the physical eye (PE) is horizontal in both cases. The shear is visualised via families of equal-distance planes. In this case the depth stretch is not noticeable. (Otherwise this would show up in the spacing.) The viewing direction of the 'mental eye' (ME) is suggested as orthogonal to the equal-depth planes. In this case the shear's slant is 18°. Notice the difference in pose in the two cases. The poses are very similar if you reckon in terms of the equal-distance planes. The shear indeed simulated a change of perspective.

has to be a change of the occlusion boundary. In that case the change of perspective involves a relative rotation of the object, in the case of the ambiguity transform it involves a shear. In many respects, though, the rotation and the shear bring about similar changes in the relief (the pattern of equal depth curves, say). Apparently the changes in pictorial relief can be interpreted as changes in 'mental perspective'. Such an interpretation is formally at least certainly possible; whether it makes psychological sense is another matter. However, we believe it to be amenable to empirical test.

Once one makes this connection it becomes clearer how these changes might be expected to depend upon the stimulus. For instance, in the case of stimulus (c) (the 'walking turtle'), the object is clearly bilaterally symmetrical, whereas the photograph has been taken from an oblique angle. One 'generic' view of the turtle would be at right angles to the plane of its support, making the 'highest point' on the turtle's shell (with respect to the support) locally frontoparallel. This is indeed the 'mental viewpoint' assumed by all four observers in the case of the task involving cross-section reproduction with horizontal replica. Such an explanation might serve to account for the similarity of the observers in resolving the ambiguity left by the depth cues. In many cases, there will be a (small) number of 'good' mental perspectives ('canonical views'), thus this general type of reasoning can also be used to forge accounts for the differences between (groups of) observers.

Such a handle suggests that one might perhaps design stimuli at will that would induce observers to resolve the ambiguity left by the depth cues in specific ways, and thus induce observers to produce specific pictorial reliefs. This would open up a novel field of enquiry.

References

Belhumeur P, Kriegman D J, Yuille A L, 1997 "The bas-relief ambiguity", in *Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Washington, DC: IEEE Computer Society Press) pp 1060 – 1066

Ernst M O, Banks M S, Bülthoff H H, 2000 "Touch can change visual slant perception" *Nature Neuroscience* **3** 69 – 73

Frisby J P, Buckley D, Bayliss F, Freeman J, 1995 "Integration of conflicting stereo and texture cues in quasi-natural viewing of a torso sculpture" *Perception* **24** Supplement, 136

Geist S, 1975 *Brancusi: The Sculpture and Drawings* (New York: Harry N Abrams)

Gibson J J, 1950 *The Perception of the Visual World* (Boston, MA: Houghton Mifflin)

Glennerster A, Rogers B J, Bradshaw M F, 1996 "Stereoscopic depth constancy depends on the subject's task" *Vision Research* **36** 3441 – 3456

Haffenden A M, Goodale M A, 2000 "Independent effects of pictorial displays on perception and action" *Vision Research* **40** 1597 – 1607

Hildebrand A, 1893/1945 *Das Problem der Form in der Bildenden Kunst* (Strassburg: Heitz) [translated by M Meyer, R M Ogden, 1945 *The Problem of Form in Painting and Sculpture* (New York: G E Stechert)]

Koenderink J J, Doorn A J van, 1997 "The generic bilinear calibration – estimation problem" *International Journal of Computer Vision* **23** 217 – 234

Koenderink J J, Doorn A J van, Kappers A M L, 1992 "Surface perception in pictures" *Perception & Psychophysics* **52** 487 – 496

Koenderink J J, Doorn A J van, Kappers A M L, 1996 "Pictorial surface attitude and local depth comparisons" *Perception & Psychophysics* **58** 163 – 173

Todd J T, Koenderink J J, Doorn A J van, Kappers A M L, 1996 "Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces" *Journal of Experimental Psychology: Human Perception and Performance* **22** 695 – 706

Appendix: The analytic form of the pictorial depth ambiguity

We derive the general analytic form of the ambiguity from a very simple and thus very general argument.

Consider two different scenes: in the first we use coordinates x, y, z; in the second x', y', z'. Here x, y and x', y' denote the dimensions in the image plane, whereas z and z' denote the depth dimension. We assume that the scenes yield the same images owing to parallel projection along the third dimension. Since we assume that the images of the scenes are the same, we have $x = x'$ and $y = y'$, but there is no similar restriction on the depth dimension—it is quite ambiguous. Let us assume that $z' = \mathrm{f}(x, y, z)$, where the function f is quite arbitrary. We suppose f to depend on all variables in sight, that is to say both depth z and position in the picture plane x, y. Notice that $z' = a + z$, where a is a constant, would mean that there is no ambiguity at all, since we measure depth up to an arbitrary offset. Thus the function f may be taken to have no constant term without any loss of generality.

Consider a plane $z = ux + vy + w$ in the first scene. It will map upon $z' = \mathrm{f}(x, y, ux + vy + w)$ in the second scene. At this point we introduce the key observation in our argument:

If it is indeed the case that planes can reliably be differentiated from curved surfaces, owing to cues such as shading and so forth, then the configuration in the second scene must also be a (possibly different) plane.

This clearly settles the form of the function $\mathrm{f}(x, y, z)$ though, it has to be linear in all variables. Apparently the equivalent configurations are related by

$$z' = a + bx + cy + dz\,. \tag{A1}$$

This argument thus reduces the class of ambiguities to (special) affinities. Notice the generality of the argument: It works for any depth cue that allows one to detect deviations from planarity.

In equation (A1) the constant a is just an offset, and thus irrelevant. The factor d can be interpreted as a stretch in depth. The factors b and c together form an arbitrary shear that involves depth but leaves the image plane invariant. It can most easily be described by considering its effect on a frontoparallel plane $z = z_0$ (say). We have $z' = a + bx + cy + dz_0 = z'_0 + bx + cy$ (where we introduce the—irrelevant—constant $z'_0 = a + dz_0$ for the sake of simplification). The line $bx + cy = 0$, $z' = z'_0$ is clearly frontoparallel (for $z' = z'_0$ denotes a frontoparallel plane). Its direction $\varphi = \arctan(c/b)$ is the orientation of the shear. The plane slants at an angle $\vartheta = \arctan\sqrt{b^2 + c^2}$. The sine of this angle, $\sin\vartheta$, is a convenient measure for the magnitude of the shear; it may assume values in the range 0 to 1.

© 2001 a Pion publication printed in Great Britain