

Local shape of pictorial relief

Jan Koenderink

Laboratory of Experimental Psychology, University of Leuven (KU Leuven), Tiensestraat 102, Box 3711, B-3000 Leuven, Belgium; and Faculteit Sociale Wetenschappen, Psychologische Functieleer, Universiteit Utrecht, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands;

e-mail: Jan.Koenderink@ppw.kuleuven.be

Andrea van Doorn

Faculteit Sociale Wetenschappen, Psychologische Functieleer, Universiteit Utrecht, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands;

e-mail: andrea.vandoorn@telfort.nl

Johan Wagemans

Laboratory of Experimental Psychology, University of Leuven (KU Leuven), Tiensestraat 102, Box 3711, B-3000 Leuven, Belgium;

e-mail: Johan.Wagemans@psy.kuleuven.be

Received 17 April 2014, in revised form 10 August 2014; published online 30 August 2014.

Abstract. How is pictorial relief represented in visual awareness? Certainly not as a “depth map,” but perhaps as a map of local surface attitudes (Koenderink & van Doorn, 1995). Here we consider the possibility that observers might instead, or concurrently, represent local surface shape, a geometrical invariant with respect to motions. Observers judge local surface shape, in a picture of a piece of sculpture, on a five-point categorical scale. Categories are cap–ridge–saddle–rut–cup–flat, where “flat” denotes the absence of shape. We find that observers readily perform such a task, with full resolution of a shape index scale (cap–ridge–saddle–rut–cup), and with excellent self-consistency over days. There exist remarkable inter-observer differences. Over a group of 10 naive observers we find that the dispersion of judgments peaks at the saddle category. There may be a relation of this finding to the history of the topic—Alberti’s (1827) omission of the saddle category in his purportedly exhaustive catalog of local surface shapes.

Keywords: shape perception, surface perception, picture perception, depth perception, shape scale

1 Introduction

1.1 Consideration involving “pictorial space”

“Pictorial space” is an aspect of visual awareness, when the awareness is in pictorial mode. It frequently happens when you look “into,” as opposed to “at,” pictures (Ames, 1925a, 1925b; Claparède, 1904; Enright, 1991; Hildebrand, 1893; Koenderink, van Doorn, & Wagemans, 2011; Pollack, 1955; Schlossberg, 1941; Schwartz, 1971), but it may also happen when you look “into” the clouds, or “into” a dirty old wall (Leonardo da Vinci, 1651). Pictorial awareness appears categorically different from more frequent modes of visual awareness. Unfortunately, it is not always obvious how to identify the nature of the awareness in experimental phenomenology. Observers monitor the mode of their present visual awareness as a level of momentary reality.

Consider some of the major differences between “generic,” and pictorial visual awareness. In daily interactions with the physical environment the eyes play an important role, even though one is not immediately aware of their functions at any moment. Most visual processes run automatically, independent of awareness. This is the major function of the eyes from the perspective of biological fitness. Theories of enactive perception (Gibson, 1950) apply. In the cases of “good looks,” or even scrutiny, visual awareness remains closely tied to the environment. The visual ego-center (“eye” for short) is experienced as being “in,” or “part of” that environment. Objects in visual awareness are in the same space as the eye. They are experienced as having backsides, even if one momentarily sees only their frontsides. That is because one may change perspective voluntarily, and reckon with others having a different perspective. Man shares this type of awareness with all other vertebrates (Spelke, 2000; Vallortigara, Chiandetti, Rugani, Sovrano, & Regolin, 2010).



Figure 1. Left: on the poster (designed by Alfred Leete, first appeared as cover of “London Opinion,” on the 5th of September 1914) Lord Kitchener points right at you, quite independent of the perspective you take. This shows vividly that pictorial space is not “connected” to the space you move in. Center: Magritte “The Schoolmaster” (1954). Notice that there is no geometrical transformation of pictorial space that will allow you to see the face. Right: illustration (“Full muscular detail”) from Rimmer’s *Art Anatomy* (1877), drawn from the imagination. This evokes the impression of a highly articulated pictorial relief. Notice how local shape varies from point to point over the surface.

These observations are crucial in understanding the pictorial mode. In pictorial mode one cannot voluntarily change the physical perspective, although the mental perspective (Koenderink, van Doorn, Kappers, & Todd, 2001) may vary. Such a change of mental perspective cannot reveal novel geometrical aspects. The eye is not in pictorial space. It is “elsewhere” in the sense that it is logically impossible for the eye to be in pictorial space. This is different from being outside a house, which one may simply re-enter. Nothing in physical space is in pictorial space, not even the picture surface if there is one. That is why Lord Kitchener (in Figure 1 left) always points right at any observer, even when they view it from an awkward angle.

In summary, pictorial space is a mental entity, without proper physical “cause.” The picture of a horse is not a physical horse, but a piece of paper covered with some arrangement of pigments. Thus pictorial objects have to be studied through methods of experimental phenomenology. Moreover, as discussed above, the geometry of pictorial space is not Euclidean. Empirical methods have to take this into account. We have developed a battery of empirical methods in the past. This project is still in a continual stage of development. At this moment our understanding of pictorial space, and pictorial reliefs, is still very incomplete.

1.2 Representation of pictorial relief: Planelets and surflets

In this study we aim to study a novel method, designed to probe properties of opaque pictorial objects. Such objects are experienced as volumetrically complete—even though they lack backsides—spatial entities, perceived through their frontal surfaces, usually denoted as “pictorial reliefs.” These objects have locations, spatial attitudes, and shapes, although these geometrical entities lack proper physical counterparts.

Since the physical eye, and the pictorial object are mutually “elsewhere,” there exists no rational way to define a “distance” between them (Koenderink et al., 2011). In some sense the eye is at “infinite” distance from any pictorial location. However, pictorial objects do have mutually different locations. This is perhaps somewhat understandable in the case of the fronto-parallel dimensions—although one might rightfully object to this relation being trivial (Lotze, 1852)—but it is certainly problematic in the case of the “third” dimension. This latter dimension is conventionally known as “depth” (Koenderink et al., 2011). Absolute depth is a non-entity, but depth differences between pictorial locations make sense. The length unit is certainly incommensurable with that applied in the fronto-parallel though. Thus one prefers to conceive of pictorial space as “2+1”, rather than three-dimensional.

On the basis of a large body of empirical data, we have formulated a geometry of pictorial space with excellent descriptive and predictive power (Koenderink & van Doorn, 2012). It is a non-Euclidean geometry of a very simple type, in that the third dimension is isotropic (Bell, 1998; Strubecker, 1941,

1942a, 1942b, 1944; Yaglom, 1979). That it is non-Euclidean should be intuitively obvious from the fact that full turns about fronto-parallel axes do not exist in this space: if one looks at a portrait in en face pose, one is forever unable to see the back of the head (and vice versa, see [Figure 1](#) center). Why? Because it was never painted in the first place! But in Euclidean geometry full turns about arbitrary axes are nothing special, hence the problem.

In a scientific investigation one desires to simplify things as much as possible, without losing the relevant aspects in the process. In geometry one of such attempts is to zoom in on local structure—so-called “differential geometry” (Coxeter, 1961).

The most local description of a narrow region on a surface is its location. This approximates the surface as a single point, for formal reasons it is known as the “zeroth order” approximation (Bell, 1998). The surface then exists as a “depth map.” The next stage in articulation is to consider the local spatial attitude. One considers the local surface as a “planelet” (Barrow, 1735). This is the “first order” approximation. To treat the globe as flat—as our ancestors did—is to stick to the first order. The next stage in articulation is to consider the local deviations from planarity. One considers the local surface as a “surflet.” This is the “second order” approximation.

Planelets are fully described by their spatial attitude, say slant and tilt. For surflets the number of possibilities multiplies. In the case of the globe, one would recognize it as spherical. The first visual round-up is due to Alberti (1436), who recognizes such surflets as “like the outside of egg shells,” “like the inside of egg shells,” “like columns,” “like the inside of reeds,” and “like a water surface.” These possibilities are qualitatively distinct, as opposed to the lower orders, where the various cases are only quantitatively different. This was generally affirmed for several centuries, till Gauss (1827) published his catalog, which added “like a horse’s saddle.” Of course, Gauss’ paper is framed in terms of mathematical formalism. Alberti thinks of shape like a continuously distributed surface quality, much like color, a surflet map.

Formally, the “quality” that one colloquially calls “shape” can be indexed through the signs of the curvatures of two mutually orthogonal sections. Generic possibilities are (Figure 2) ++ (cap, that is “like the outside of egg shells”), +− (saddle, that is “like a horse’s saddle”) and −− (cup, that is “like the inside of egg shells”), with transitional forms +0 (ridge, that is “like columns”), and −0 (ruts, that is “like the inside of reeds”). A non-generic case is 00 (flat, that is “like a water surface”), which is actually a quantitative, rather than qualitative distinction (see Appendix at <http://i-perception.perceptionweb.com/journal/1/volume/5/article/i0659>). The generic surflets occur over areas, whereas the transitional forms occur on certain special curves bounding such areas (see Appendix).

1.3 Aim of the present study

It is not known how reliefs are represented in human vision. Obvious possibilities are in terms of depth maps, planelet maps, surflet maps, and so forth. [Figure 3](#) illustrates some of these possibilities. A surface represented by a finite number of points, or depth values, is fully determined. A surface represented through a planelet map is somewhat ambiguous, in the sense that its overall depth is indeterminate. A surface represented through a surflet map is even more ambiguous, in the sense that both its overall depth and overall slant are indeterminate. Whatever the representation, observers will be able to judge other quantities, up to the ambiguities mentioned above. However, such quantities have

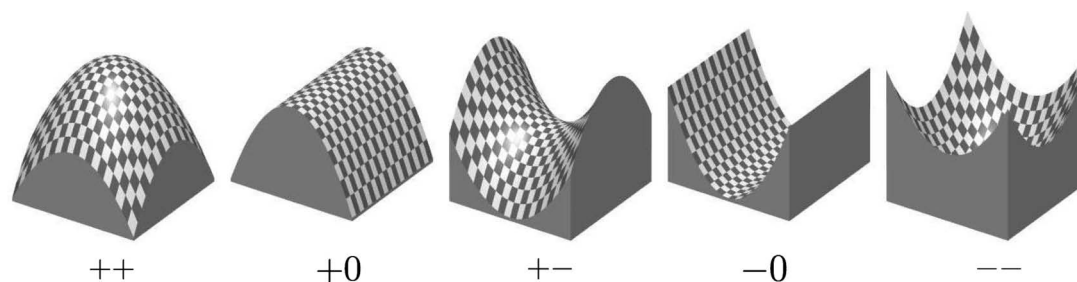


Figure 2. Examples of the generic surflets. From left to right: CAP (++, “like the outside of egg shells”), RIDGE (+0, “like columns”), SADDLE (+−, “like a horse’s saddle”), RUT (−0, “like the inside of reeds”), and CUP (−−, “like the inside of egg shells”). Ridge and rut are transitional forms that occur on the common boundaries of areas of saddles, and areas of cups or caps.

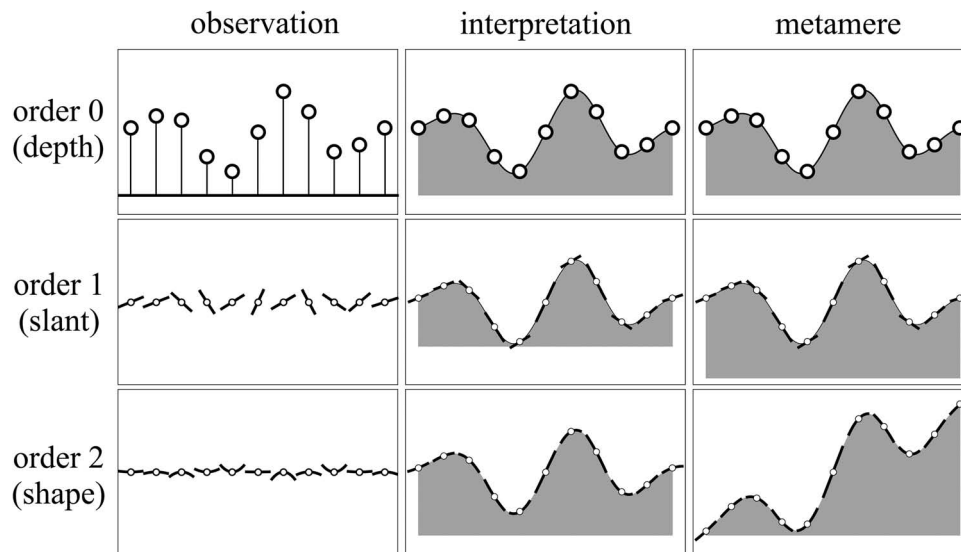


Figure 3. In the left column we show likely relief representations: depth maps (zeroth order), planelet maps (first order), and surflet maps (second order). Of course, even higher order representations are possible! In the second column we show a possible awareness, a smooth relief, based on the finite representation. In the third column we show another—equally possible—interpretation. Only the absolute depth map (order 0) is fully determinate. In the planelet representation (order 1) absolute depth is indeterminate, in the surflet representation (order 2) both absolute depth and overall attitude are indeterminate. These are useful thought models; reality might well be a mixture of diverse, incomplete representations.

to be derived from the basic representation, thus they cannot be more precisely determined than that. Empirically, we know that observers are unable to judge absolute depth, but can judge relative depths (Koenderink & van Doorn, 1995). They can also judge surface attitude (Koenderink, van Doorn, & Kappers, 1992) and shape (this paper).

For the sake of clarity, Figure 3 shows a one-dimensional “image.” The vertical direction represents the depth domain. Although this suffices to make basic conceptual points, it fails to show an important complication that occurs in higher dimensional cases. Especially the two-dimensional (“image”) case is important here. A planelet map does not necessarily “mesh” into an integral surface. There exists an integrability constraint, namely the planelet map should be a gradient field, or be “curl-free” (a curl-free field has no vortices; Spivak, 1999; see Figure 4.) This can be empirically tested (Koenderink et al., 1992), and is one of the most important indicators for the nature of the relief representation. A similar condition has to be met in the case of surflets. In the case of Euclidean geometry, the surflet map has to satisfy the so-called “Codazzi–Mainardi equations” (Codazzi, 1868,

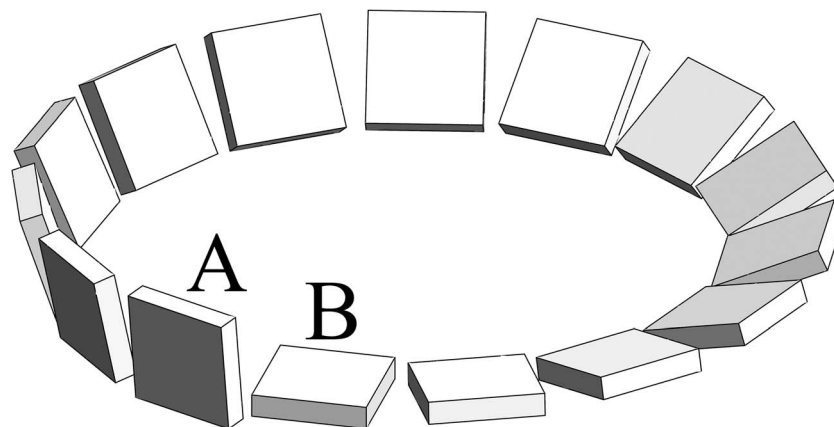


Figure 4. An example of a circular string of planelets, taken from a planelet map that fails to “mesh.” Notice that the adjacent planelets A and B are far from a coplanar condition. This happens when a planelet map is not a gradient field, or “curl free.” In such a case there does not exist an integral surface, thus not a depth map.

[1869](#); Mainardi, [1856](#); Spivak, [1999](#)). This is—again—up to empirical test, at least in principle. In the case of the geometry of pictorial space, the formal condition is slightly simpler. Details are available in textbooks (e.g., Sachs, [1990](#)).

We have investigated the meshing of observed planelet maps, and found that the integrability condition is met within the empirical spread (Koenderink et al., [1992](#)). We have also been able to show that empirically planelet maps are more precise than relative depth maps (Koenderink & van Doorn, [1995](#)). Might surflet maps be sufficient to “explain” planelet maps? Yes, that is to say, up to an overall ambiguity of spatial attitude. This is perhaps okay, since we often find idiosyncratic overall attitude variations in empirical planelet maps. A formal argument in favor of a surflet representation is that shape is an invariant over arbitrary motions. Translation will affect location, and rotations will affect spatial attitude, but shapes remain unaffected. This might be an advantage where mental models are concerned.

Although extremely important from a conceptual perspective, it will take much effort to make progress towards an answer. In this paper we only attempt a first step, in that we construct methods to sample surflet maps.

1.4 Proposed method to sample surflet representations

Consider some generic relief. An example would be the pictorial relief evoked by a photograph, drawing, or painting of some articulate object, like a sculptural representation of a human torso. The example shown in [Figure 1](#) (right) provides an instance. In such a relief the local shape varies continuously from location to location. A point on such a surface is at some depth, the planelet at the point has some spatial attitude, and the surflet at the point has some shape. The shape is the local variation of the planelets, the planelet the local variation of the depth. A really large neighborhood will reveal a variation of surflets, a very restricted neighborhood may show only irrelevant textural variation, thus the scale, that is the size of the relevant region of interest, is a crucial parameter (see [Appendix](#)).

In sampling a surface one uses a grid of sample points, the grid spacing setting the scale at which the sampling proceeds. In a typical experiment one may consider a grid of a hundred to a thousand points, thus the scale is limited to roughly a few up to 10 percent of the width of the pictorial object. Of course, here the “sampling” is a property of the experimental paradigm. It is quite unclear how much of the image is “used” by the observer to arrive at a “local” judgment. Moreover, this is not something that we can ascertain from the present investigation. On the one hand, the obvious shape variations over the depicted object suggest that the area available for a “measurement” is rather limited. On the other hand, it is intuitively clear that observers would be at a loss when you would mask all of the image except this crucial area. This is conceptually a very difficult problem. We have explained the formal issues in the [Appendix](#). One finds that there is always a fairly well-defined “footprint” that is required, although much larger areas, perhaps even the whole image, are probably needed to provide the necessary context. For instance, if the major cue is shading, the context might be the direction of illumination. Notice that an estimate of the direction of illumination requires an understanding of the shape map. Thus the issue is a very involved one. We leave it for future investigation.

In the simplest case one indicates a location on the relief and asks the observer to specify the nature of the shape of the relief in the immediate neighborhood of that location. For instance, it may be seen as “convex” (like the outside of egg shells), and so forth. In order to turn this simple task into a formal method of “measurement,” one needs to provide a formal scale. This is the nature of the method we propose here.

Technically, one uses the ratio of extremal sectional curvatures to define a size-independent shape measure. This “shape index” (Koenderink, [1990](#); Koenderink & van Doorn, [1992](#)) is defined on a closed linear interval, with the cap and cup at its extremities, the symmetrical saddle at the center. There exists a natural symmetry about the center, since shapes situated at mutually symmetric points are related as a plaster cast with its mold. Thus, for instance, cup and cap are related in this way, whereas the symmetrical saddle is congruent to its own mold.

Although this is a continuous scale, it can be coarse-grained into natural categories. This is done by noting the signs of the extremal curvatures. Notice that one needs some convention, say curvature reckoned positive for convexity, negative for concavity. This yields the ridges and ruts as natural anchors between the center and the end-points of the scale. Then the cap–ridge–saddle–rut–cup sequence appears as a natural set of shape categories. It is the scale that was illustrated above in [Figure 2](#).

This should not be interpreted such that “ridge” (for instance) implies that one principal curvature would be identically zero. Indeed, that would make no sense, because it would indicate a *singular case*, thus an *empty category*. One should understand “ridge” as a central item in its category. Some things are perceptually more “ridge” than either “cap” or “saddle.” Such things are immediately obvious to the observer in the actual task, but we noticed that it leads to formal misunderstandings by colleagues who merely conceptually *think about*, rather than *perform* the task.

In this first investigation of surflet maps, or “shape landscapes,” we ask observers to categorize local shape on this categorical scale. This scale is augmented with a “flat” category, required because observers cannot always detect a curvature. Planarity is not a shape, no shape index can be assigned to it. It is perhaps most appropriate to say that flat is *any shape you want*. In formal geometry, planar points occur with probability zero. However, in the case of uncertainty, there is a finite probability that the curvatures are not significantly different from zero. Then it becomes meaningful to consider a “flat” category. Of course, this does not imply that “flat” is a proper shape (see [Appendix](#)).

We then study the consistency of such judgments for given observers over repeated sessions, as well as the degree of agreement between observers for given locations. Both intra-observer consistency and inter-observer agreement can be studied as a function of shape, for instance, by relating them to the median over numerous judgments. Notice that this does not involve any notion of “ground truth.” Thus the study is singularly limited to pictorial space.

2 Methods

2.1 The stimulus

In this study we use a photograph of a piece of sculpture by Andrew Smith (working as professional sculptor since 1989; Smith, [2014](#)). It is simply labeled “Reclining Nude—Portland stone” ([Figure 5](#)). We do not know the exact size of the piece, nor the lighting set-up for the photograph, distance or focal length, etc. Portland stone is a limestone from the Tithonian stage of the Jurassic period quarried on the Isle of Portland, Dorset. It has found many uses in the UK, e.g., St Paul’s Cathedral and Buckingham Palace are constructed from it. It is an oolitic limestone that is so well cemented that it can be readily worked, and is quite suited to larger pieces of sculpture. What is important here is that it has slight texture, somewhat apparent in the image. Its surface scattering is diffuse, not all that different from Lambertian.

Most of the visible relief in the photograph is revealed through conventional “shading” (Baxandall, [1995](#); Horn & Brooks, [1989](#); Lambert, [1760](#); Metzger, [1975](#); Ramachandran, [1988](#)). The pose is such that the occluding contour is rather revealing. Occlusions (T-junctions—ending contour pairs) are of minor importance (waist, shoulders). Parts where the relief emerges from the block (legs, arm, part of the thorax) are clearly marked, and not easily confused with occluding contours. A few singular curves (cleft between the legs, cleft between the buttocks) are easily read, and not likely to be confused with occluding contours either. Thus this is a very revealing photograph of a generic (Koenderink, [1990](#); Porteous, [1994](#); Thom, [1972](#)), predominantly smooth relief, which renders it eminently suitable to our purpose.

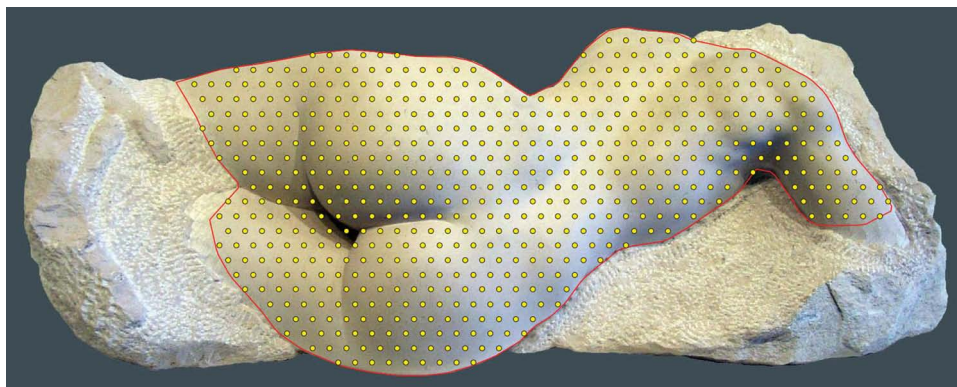


Figure 5. The stimulus is a photograph of a sculpture “Reclining Nude,” executed in Portland stone by Andrew Smith (<http://www.assculpture.co.uk/reclining%20nude.html>). The red line is a contour that defines the region of interest, the yellow dots show the 600 fiducial locations used to sample apparent surface shape.

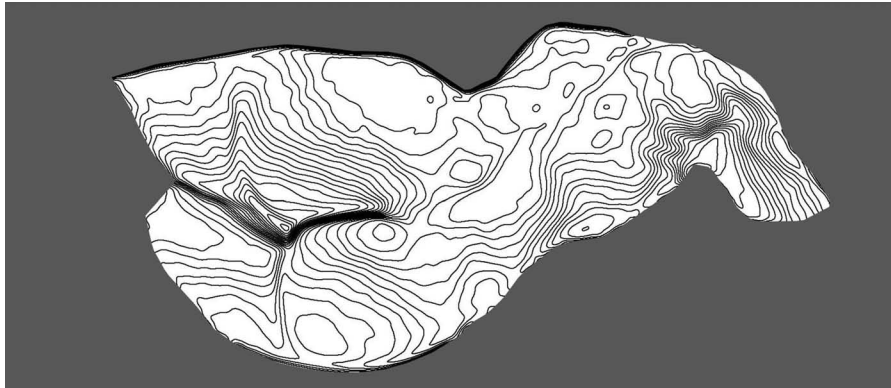


Figure 6. Isophotes in the photograph of the “Reclining Nude.” In order to find them we first established the scale through Gaussian blurring by about half the edge length of the triangulation, followed by a “posterization,” again followed by an “edge finding” operation. The whole procedure is easily finished in Photoshop. This pattern defines the shading cue. One readily identifies the light direction (remember that the buttocks are almost spherical), the cylinder axes (legs, spinal area), and a variety of cylindrical points (most of the extrema, and all of the saddles in the dorsal thorax). Also notice the effect of the cast shadow of the upper buttock on the lower leg and the reflex at the lower left contour.

The piece is evidently illuminated from top right (Figures 6, 7). The (minor) cast shadows reveal three rather directional sources, from only slightly different directions. Thus most of the relief can be regarded as being illuminated through a single, directional, effective source. There are traces from diffuse reflexes (e.g., at the bottom left). Some “shading” is actually due to vignetting (cleft between the legs, arm pit).

The “reclining nude,” predominantly female, is a conventional topic in Western art (Clark, 1956; Rogers, 1969). Both dorsal and ventral views are common. The artistic interest is in the composition of overall pose and local anatomical detail. Different from common misconception, a piece like this is to be understood as a sophisticated abstract work. The nude is not a subject of art, but a form of art. Although major body parts can easily be named, which is quite useful in discussing the relief, this does not at all imply that one would know the relief without even looking at it. In fact, most of the important details could hardly be named by the naive observer. Who actually remembers the location, and shape of the infraspinatus, or teres minor and major? The anatomical knowledge of most observ-

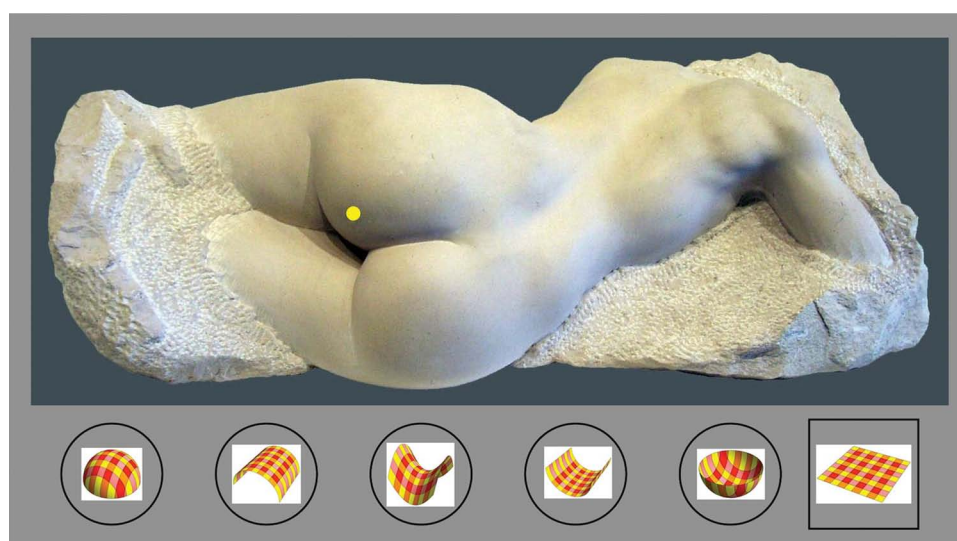


Figure 7. An impression of the screen layout. The “flat” category is added by way of a “don’t know” escape, and does not properly belong to the shape index scale. In fact, any shape, when sufficiently attenuated will become “flat.” One might say that flat is to shape, as black is to hue. The yellow dot indicates the fiducial location that identifies the trial. (For the sake of clarity, the size is enlarged in print.) The observer responds by clicking the appropriate category with the mouse. In this case the “cap” category will appear appropriate to most observers.

ers is very limited anyway, and surface articulations easily as prominent as the nose are easily missed by non-professionals. But notice that although most people certainly could point out a nose in a face, they would be hard put to indicate its boundaries—even less so with a cheek. Such “knowledge” is very poor indeed. Even a medical training is only of limited use, because “artistic anatomy” is a topic by itself. In artistic anatomy one routinely merges objects that would be distinct to the medical person. The artistic taxonomy echoes the “Gestalt understanding” of the human form, not the anatomical one in the scientific sense (Bammes, 2002; Hamm, 1983; Hatton, 1910; Hogarth, 1958).

The bottom line is that the relief is largely an abstract entity to most observers. One is aware of an array of mutually interconnected “objects” that have a fleeting existence and constantly reorganize during continued scrutiny. Their lifetime is only a moment of visual awareness.

2.2 Set-up

The stimulus was presented on a DELL U2410f monitor, a $1,920 \times 1,200$ pixels liquid crystal display (LCD) screen, in a darkened room. The viewing distance was 78 cm. The stimulus filled the width of the screen. Above and below were free areas used for user interaction. In the upper area were placed a progress bar, a record of time spent in the session, and an analog clock showing time of day. The lower area was used for the actual user interface (see Figure 7).

In all cases, viewing was binocular, using open view, possibly using one’s regular correction. Interaction was by way of the mouse (see below). At each trial a mark (a filled dot) indicated a fiducial location, and the observer had to judge the local shape by clicking one of the categories presented just below the picture (Figure 7). Observers were instructed to approach the task very seriously and use all the time they needed to do that. Responses were collected as XML files and processed later, off-line. We recorded both response times and actual responses, although the response time data were hardly used. Trial sequence was randomized in all cases.

2.3 Observers

We used two distinct groups of observers. The first group of 10 observers was fully naive with respect to the aim of the experiment. Both genders were present in roughly equal proportions, all were aged in their 20s to 30s. Most had prior experience with various types of visual experiments. Each of these observers performed a single session. The second group of observers were the authors AD (female, in the 60s), JK (male, in the 70s), and JW (male, in the 50s). All are experienced with various psychophysical procedures, but effectively naive in the present task. They have some basic understanding of geometry, and of artistic vision, but are not active as mathematicians, nor as visual artists. Each repeated the session three times, at different dates. We discuss these groups separately, since different methods of analysis apply.

3 Results

3.1 First group

The first group comprises 10 observers. Our first concern is to check to what degree they yield mutually similar results, and whether it is possible to detect obvious outliers from a global perspective. A first global check is to consider the response times. Although observers were instructed to go at their own speed, the duration of the session was long enough that they no doubt tried to economize. We find that the median response times vary from 2.1 to 3.7 s, with an interquartile range (i.q.r.) of 0.38 s. The interquartile ranges of all observers overlap, thus in that sense there are no obvious outliers.

Another global check is to consider the distribution over the categories over all locations. Here we find obvious differences (Figure 8). There can be little doubt that there exist very significant inter-observer differences.

Of course, the distribution over locations is crucial. As a first check one may consider whether responses are consistent per location. A simple way to do this is to compare observers pairwise. For a given pair we consider only instances where both observers did not report “flat.” Since the responses are labeled by location, we may then simply consider a rank order correlation. We used the Kendall tau. It turns out that the observers correlate very significantly. The Kendall tau has a standard deviation of about 0.033. The correlations have a median value of 0.58 (i.q.r.: 0.49–0.66). A cursory glance at the distribution over all pairs (Figure 9) reveals one observer as an obvious outlier. It is observer #10, the one that is at the bottom of the sorted list shown in Figure 8.

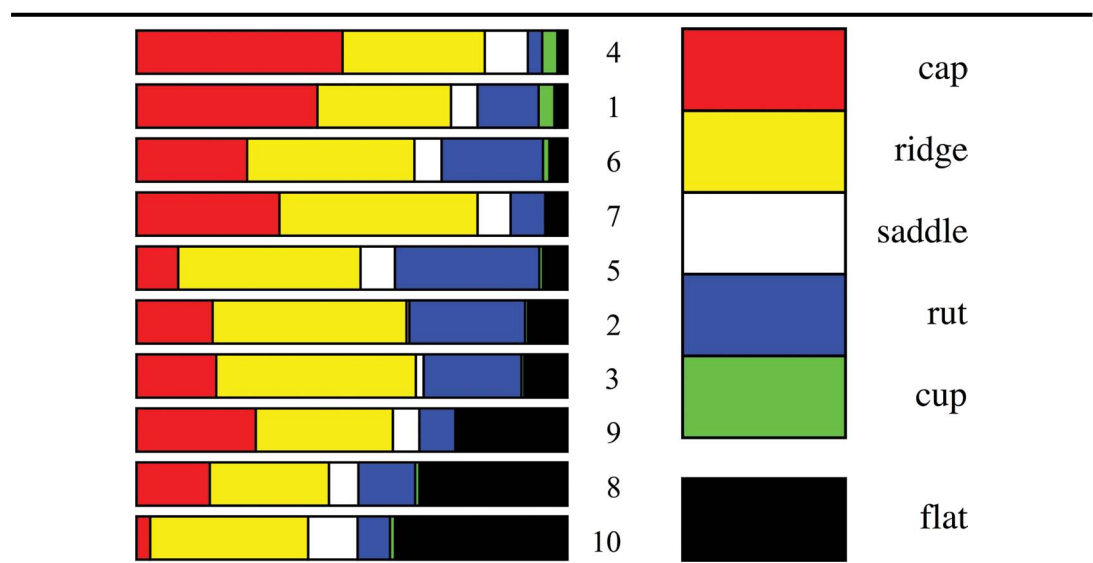


Figure 8. At left the distribution of the responses over the categories is shown for the first group of observers. Observers #1– #10 have been sorted by increasing fraction of “flat responses.” Here, as in all subsequent figures, we use the canonical color code shown at right for the shape categories (Koenderink & van Doorn, 1992): red—cup, yellow—ridge, white—saddle, blue—rut, green—cup. We have indicated the flat category with black, which is colorless, thus indicating the lack of shape. Notice that opponent hues are used to encode opposite (as shape and mold) shapes, this is why the saddle is shown as white.

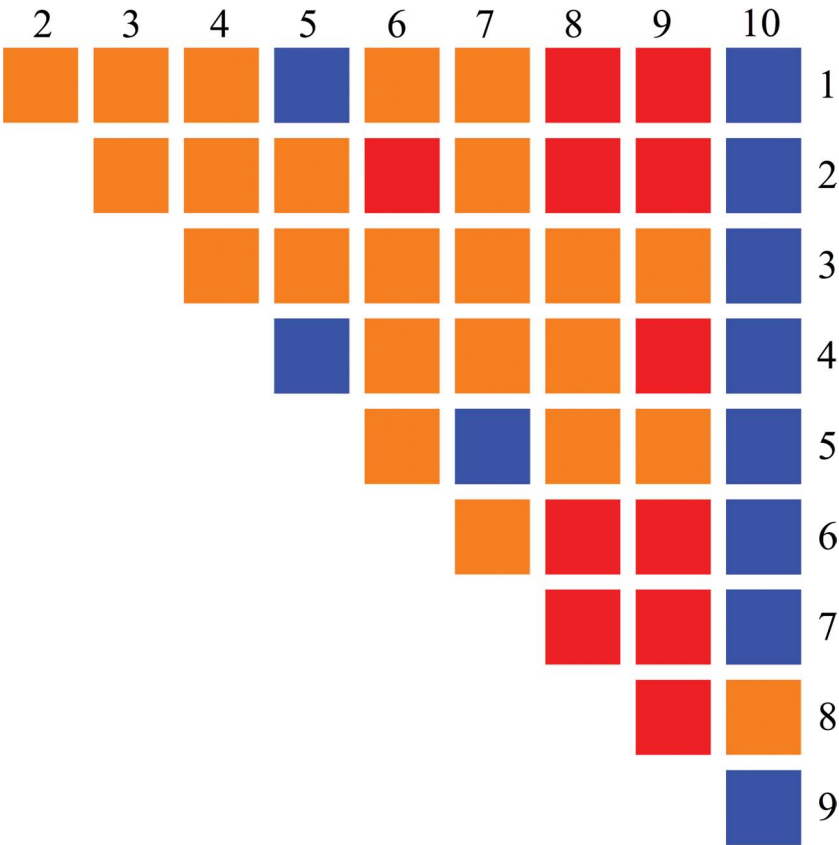


Figure 9. Here we consider the rank correlations (Kendall’s tau) for all pairs of observers. The colors indicate: orange: within the interquartile interval; red: above the 0.75 quartile; blue: below the 0.25 quartile. Notice that observer #10 is an obvious outlier.

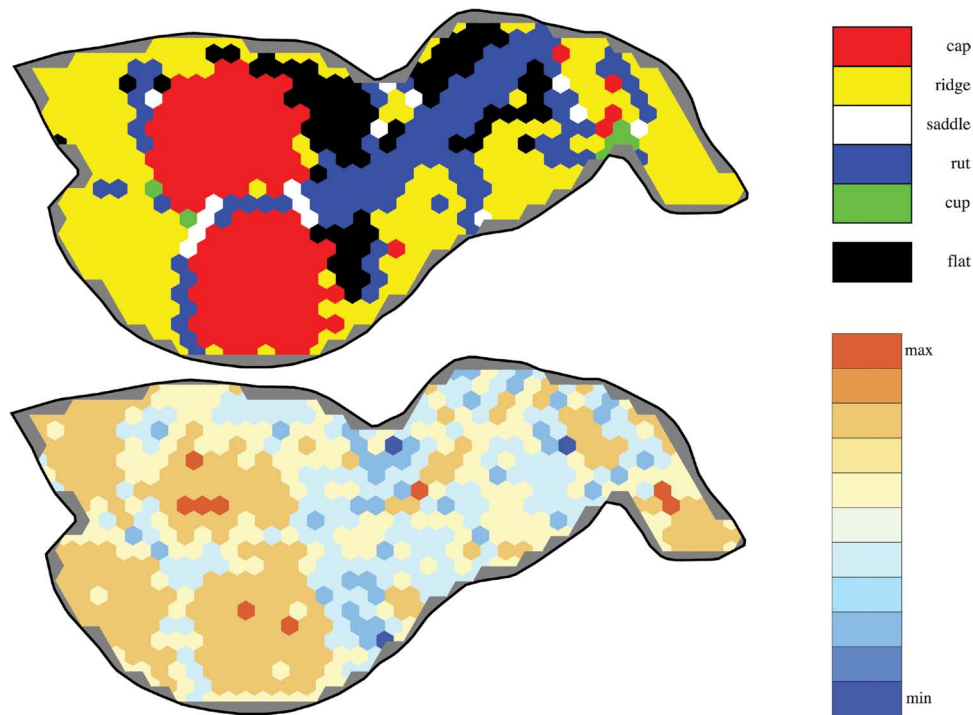


Figure 10. At top we show the consensus (by majority vote) for observers #1–#9, using the color code as in figure 8. At the bottom we show the distribution of mutual concordance, normalized on the temperature scale from blue (least) to red (best), shown at right.

Although it is clear that observer #10 is an outlier and perhaps should be ignored when considering overall trends, it is of obvious interest to see in which respects this observer differs from the others. In order to find out, we need to consider the actual spatial distribution of responses. First, we study the overall consensus for all observers except #10 ([Figure 10](#)). The concordance is quite good; it is the least in the area of the back above the waist, with the exception of the spinal rut.

In comparison, the distribution for observer #10 is quite exceptional ([Figure 11](#)). This observer rates about 40 percent of the area as flat, and most of the remainder as ridge. A few cups and caps appear randomly scattered. A saddle area seems somewhat coherent, but apparently coincides with the points that the other observers frequently rated as “rut.” We conclude that this observer is hardly aware of a consistent pictorial relief at all. One might say that this observer experiences the stimulus largely veridically, that is to say, as a flat picture. This is the reason to omit the data of observer #10 from our overall assessment.

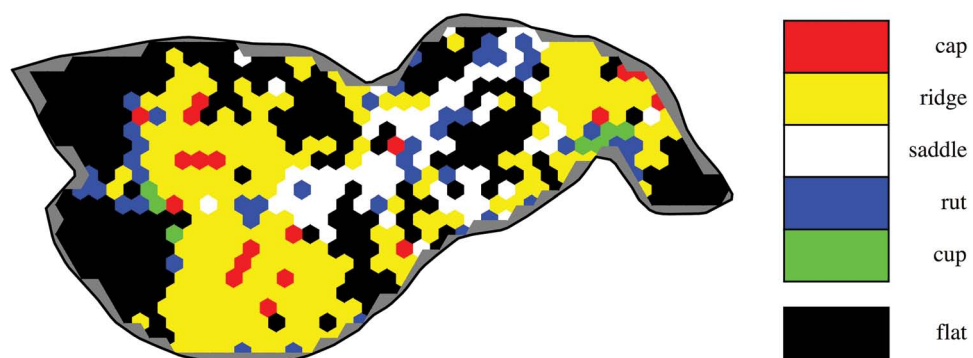


Figure 11. The distribution of shape responses for observer #10. Notice that much of the area is judged flat, most of the remainder as convex cylindrical (ridge).

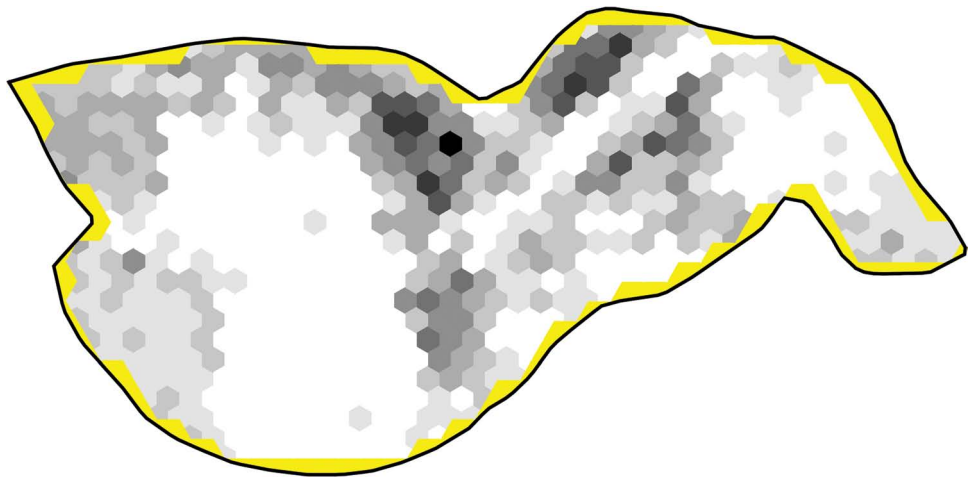


Figure 12. The distribution of “flat” responses, averaged over the observers #1–#9. Gray tone represents the frequency of flat ratings. White means never, black means rated flat by all observers.

In [Figure 12](#) we plot the distribution of “flat” responses, averaged over the nine remaining observers. One notices an obvious pattern: the shape is well defined in the buttocks, spinal, and exposed shoulder areas; it is ill defined in the waist and the scapular areas.

Apart from the distribution of responses over locations, it is of much interest to study the distribution of responses over the categories. We summarize our findings in [Figures 13](#) and [14](#). These figures are based on the pooled responses of observers #1–#9.

There are a few difficulties that need to be overcome in this study. A trivial, but serious problem is the very unequal distribution over the responded categories. Consider the pie-chart (over observers #1–#9) shown in [Figure 13](#). This problem is most readily solved by normalizing over this unequal distribution. The only drawback is that the distribution for the cups will be far less precise than that over the ridges, and so forth. This is not too problematic. Below we will point it out where necessary.

More of a conceptual problem is that there does not exist anything like a notion of “ground truth.” The only solution appears to be to study the distribution of responses factored with respect to the average, or median response per location. We use the median for reasons of robustness. We obtain a set of responses for each median response; each set is normalized. The result is the distribution shown in [Figure 14](#). The distribution is roughly concentrated about the main diagonal—such would be the “ideal” result—but there is a marked dispersion, especially in the saddle region.

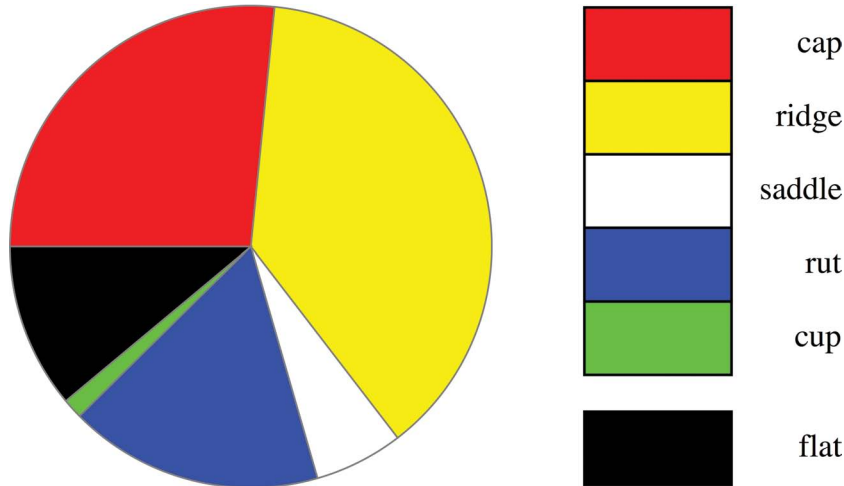


Figure 13. Pie chart of the distribution of categorical responses pooled over observers #1–#9 of group one.

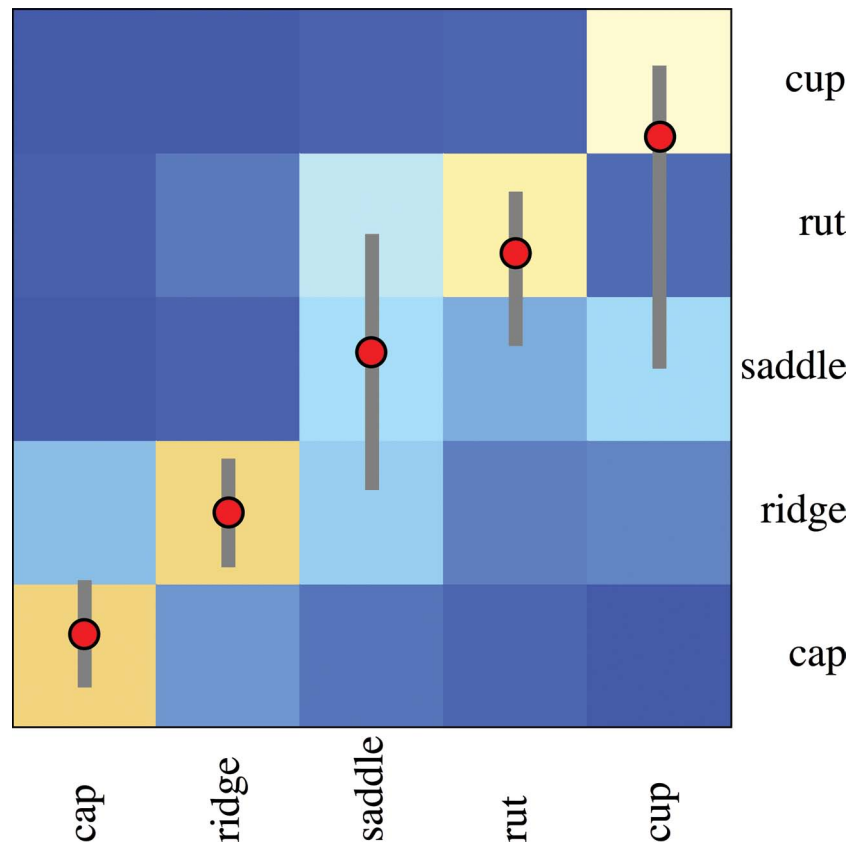


Figure 14. The distribution of responses over categories, as a function of the median category response. These are the pooled results for the observers #1–#9. The distributions have been normalized per median category, that is to say, column wise. This normalization serves to remove the effect of the (very) different frequency of the categorical responses. The extremes are the 2,211 ridge versus the 10 cup responses. We omitted the flat responses. The colors are from a temperature scale, from blue (low density) to red (high density), as in [figure 10](#). The red dots indicate the medians, the gray bars the interquartile ranges, of the distribution over a column. The cup category is based on so few instances that one should perhaps not attach too much importance on it. The width of the distribution at the saddle category is significantly larger than that of its ridge and rut neighbors.

From [Figure 14](#) we conclude that observers are not able to use the categorical scale with ultimate precision. They are quite precise in the cap (i.q.r.: 0.65 of 5), ridge (i.q.r.: 0.66), and rut (i.q.r.: 0.98) regions, but more variable in the saddle region (i.q.r.: 1.68). Because of the scarcity of instances we leave the cup region (i.q.r.: 2.01, but very variable) out of this discussion. An overall conclusion might be that observers are quite variable in the saddle region, where the uncertainty amounts to almost half the total scale (40%).

3.2 Second group

The second group of observers consisted of the authors. They did the full session three times, on different dates. Thus we gain the advantage of looking at intra-observer variations. As a first overall check we ran the same analysis as for the first group on all sessions, thus treating the repeated sessions as due to different persons. We find results that do not appear to differ from those of group one. A first look at the data involves the spatial distribution of categorical responses over all observers and all sessions. It is presented in [Figure 15](#).

In [Figure 15](#) we spot both obvious differences as well as obvious similarities. These are perhaps more easily studied in the overall consensus, and the distribution of mutual correspondences ([Figure 16](#)).

In [Figure 16](#) we see that the consensus and mutual concordance are quite similar to those of the group of fully naive observers. Apparently these groups are largely comparable. The median rank correlations over the three sessions of an observer are 0.78 (AD), 0.81 (JK), and 0.77 (JW). Over all observers and all sessions the median rank correlation is 0.65 (i.q.r.: 0.60–0.73), thus the repeated sessions per observer are more similar than the observers are like each other. Apparently there exist

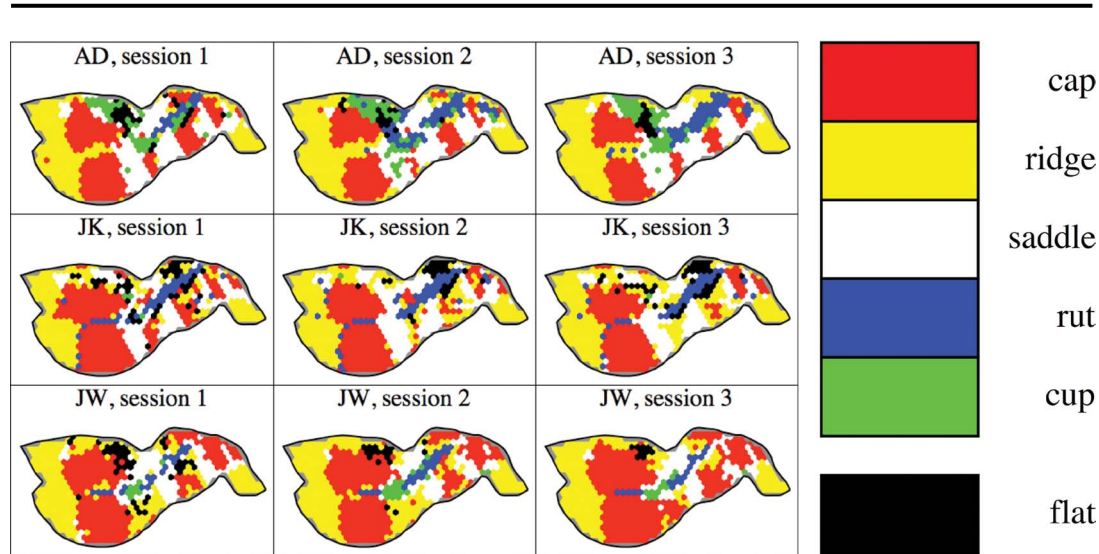


Figure 15. The distribution of categorical responses over all sessions for the three observers of group two.

individual differences. Perhaps not unexpectedly, the observers of group two are—as a group—more coherent than those of group one. This induced us to do the final analysis we did for group one once over again, but only for the sessions of a single observer (necessarily of group two). The result is shown in [Figure 17](#).

The result is striking to the extent that the observers of group two manage to retain their categorical judgments over time. The only exceptions are near the cups-extreme, where the data are rather uncertain. Apparently categorical judgments of shape are possible, although there is only moderate consensus (median rank correlation: 0.65) over observers.

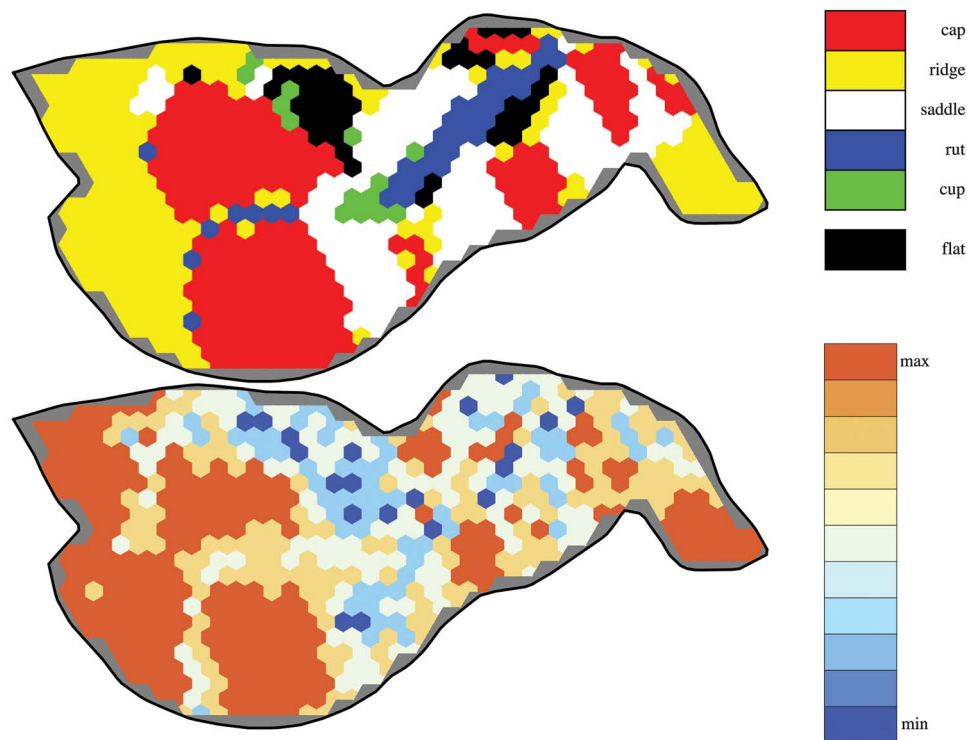


Figure 16. At top we show the consensus (by majority vote) for all session of the observers of group two. At the bottom we show the distribution of mutual concordance, normalized on the temperature scale from blue (least) to red (best).

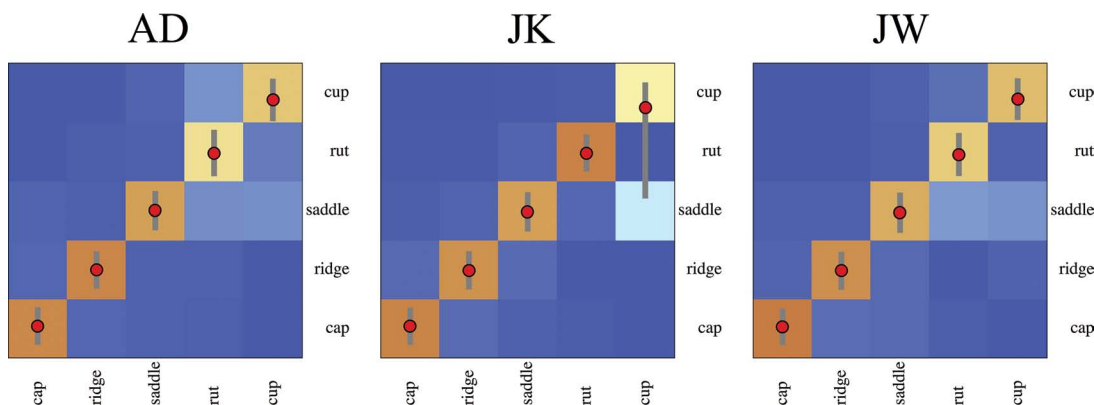


Figure 17. The distribution of responses over categories, as a function of the median category response for the sessions of each observer of group two (compare with [Figure 14](#)).

4 Conclusions

Given the wide-ranging scope of the problem our conclusions must necessarily be of a preliminary nature. However, given the fact that essentially no quantitative researches have been published thus far, we trust that they are of some interest.

Perhaps the major drawback of this study is the fact that we used only a single stimulus. For a problem like this there is no “ground truth,” thus, in the conventional sense, “no stimulus” at all. A counter-measure might be to use a large set of mutually very different pictures. We have not done this, but consider it still reasonable to draw a few interesting preliminary conclusions.

We find that observers will readily judge the local shape of a pictorial relief on a five-point categorical shape index scale (see below), needing only a second or so for a “good look,” and with a spatial resolution that is at least some five percent of the picture width. At least, this holds true for our stimulus, where the major cues to pictorial shape are contour and shading, and the rough overall nature of the pictorial object is not in question.

Notice that we do not count “flat” as a category here. From a formal, geometrical point of view flat points may be ignored, because they do not generically occur on unconstrained surfaces. From an empirical point of view one expects any of the categories to evoke “flat” responses when the curvatures are subliminal for an observer. Thus “flat” is not a true category of shape, but is defined through a quantitative criterion. Flat is related to the shape-scale in a similar way as black is related to the hue-scale. In a study like this one might perhaps use the flat responses to gain some insight into the threshold for curvedness. Due to the scarcity of instances we have refrained from doing so.

The “precision” of the shape inferences is hard to describe. We draw the tentative conclusion that good observers can resolve the five-point categorical shape index scale easily, but that different observers may well disagree by up to about half of the total scale extent. This implies a number of rather important conceptual issues.

One issue is that it makes hardly any sense to average over observers. Such averages may indeed yield a population measure, but it severely misrepresents the abilities of individuals. Averages are perhaps of interest for industry or government work, but rather less so to vision science.

Another issue is that “the” category at which one defines the spread is already a non-entity in the sense that it does not exist as an objective stimulus parameter. In this type of “experimental phenomenology” one has no option but to study the nature of the response in terms of properties of itself. This makes sense, because phenomenological research addresses the internal relations between aspects of awareness, rather than the relation of aspects of awareness to given objective stimulus parameters. The advantage is that experimental phenomenology addresses qualities and meanings, whereas psychometric studies cannot go beyond objective—thus experientially devoid of meaning and quality—properties. Both approaches are necessary and important, albeit in mutually complementary ways.

A single observer is able to use the five-point shape index scale reliably over extended periods (days). We find this not too surprising, because the scale is a natural one, based on the sign of sectional curvature. All one needs to be able to do is to judge convexity/concavity in a few directions. It is hard to image that one would not be able to do this, except when the local relief appeared flat, as it occasion-

ally does. Seeing curvature intuitively implies seeing its sign, an intuition that is by no means conclusive. Thus, seeing something as curved (or “non-flat”) seems to imply that one will be able to use the scale. On the other hand, the ability to interpolate the scale would involve abilities of a quantitative nature, such as the ability to judge the foreshortening with respect to both direction and magnitude. This puts the continuous shape-index scale in an entirely different ball park.

Why are observers so obviously different? The issue may well be the absence of a proper “physical cause” of the awareness. For if all observers strive to attain a common physical target one indeed expects them to become similar. For instance, if all observers are instructed to shoot at the same physical target in front of them, for “target practice,” one would be surprised to find major systematic misses. The case of pictures is very different. Leonardo saw medieval battles in a dirty old wall, but what would a modern observer see in the same situation? It could be virtually anything, but a medieval battle scene would be highly unlikely.

A conclusion from the previous observation might well be that it will be of much interest to vary the cue-content of the stimulus pictures parametrically in this task. Thus the method may well become a tool to study pictorial perception per se.

What about observer #10 of group one, the obvious outlier? What seems to be the case here is that this observer did not experience pictorial relief, but experienced the screen of the monitor much like it was, which is flat. The observer apparently did not look into, but at the screen. Of course, it will be very hard to make such a guess objective. One method to do so is to test the person with a battery of methods and pictures, which would imply a major undertaking. Such an effort may become worth it when one had a sizable group, say at least half a dozen, of such observers. We believe it very likely that human observers that lack the usual “pictorial mode” exist (Koenderink et al., 2011). In fact, we believe to have met quite a few of them as “outliers” in a variety of experiments over the years. One encounters them not just among naive observers, but also among vision scientists. In the latter case one may discuss the nature of their awareness with the observers. We have met with a variety of interesting cases, like inability to become visually aware—as different from reasoning out—of concavities (van Doorn, Koenderink, Todd, & Wagemans, 2012).

As we mentioned in the introduction, in 1436 Alberti came up with a taxonomy of local surface shapes that was complete except for the omission of saddles. Moreover, it took till the work of Gauss (of 1827!) before this omission was amended. How was this possible? After all, most of the comments on Alberti stem from visual artists, or intellectuals interested in the visual arts. How come no one noticed the omission a century before Gauss? Our study fails to yield a complete answer. We find that our observers readily categorize local surface shapes as belonging to the saddle category. Moreover, individual observers do not fluctuate in their response to saddles over time. However, when one compares a group of observers, one finds that the responses fluctuate much more heavily near the saddle category than in other parts of the scale. Apparently there is something special the case with the saddle category. It is unlikely to be a prior, because saddles are actually more frequent in nature than caps or cups (Koenderink & van Doorn, 2003; Lillholm & Griffin, 2007).

In summary, human observers readily categorize local pictorial relief in terms of cap, ridge, saddle, rut, cup, and flat categories. They do this consistently, but the inter-observer variability is very significant. Although all observers globally “see the same object,” there are considerable differences in detail.

Acknowledgments This work was supported by the Methusalem program by the Flemish Government (METH/08/02), awarded to Johan Wagemans. We would like to acknowledge administrative support by Agna Marien and technical support by Rudy Dekeerschieter.

References

- Alberti, L. B. (1972, original 1436). *Della Pittura*. (In English: *On Painting*. Harmondsworth: Penguin Classics).
- Ames, A. Jr. (1925a). The illusion of depth from single pictures. *Journal of the Optical Society of America*, 10, 137–148. doi:10.1364/JOSA.10.000137
- Ames, A. Jr. (1925b). Depth in pictorial art. *The Art Bulletin*, 8, 4–24. doi:10.1068/i0585
- Bammes, G. (2002). *Die Gestalt des Menschen*. Stuttgart: Urania. doi:10.1007/s12064-002-0015-6
- Barrow, I. (1735). *Geometrical lectures*. London: Printed for Stephen Austen, at the Angel and Bible in St. Paul’s Church-Yard. doi:10.1007/978-3-642-29163-0

-
- Baxandall, M. D. K. (1995). *Shadows and enlightenment*. New Haven, CT: Yale University Press. [doi:10.1364/JOSAA.18.001804](https://doi.org/10.1364/JOSAA.18.001804)
- Bell, J. L. (1998). *A primer of infinitesimal analysis*. Cambridge: Cambridge University Press. [doi:10.4236/jmp.2011.211172](https://doi.org/10.4236/jmp.2011.211172)
- Claparède, E. (1904). Stéréoscopie monoculaire paradoxale. *Annales d'Oculistique*, 132, 465–466. [doi:10.1080/00033795300200243](https://doi.org/10.1080/00033795300200243)
- Clark, K. (1956). *The nude: A study in ideal form*. Bollingen Series 35.2. New York, NY: Pantheon Books.
- Codazzi, D. (1868/1869). Sulle coordinate curvilinee d'una superficie dello spazio. *Annali di Matematica Pura ed Applicata*, 2, 101–119. [doi:10.1007/978-3-319-00690-1](https://doi.org/10.1007/978-3-319-00690-1)
- Coxeter, H. S. M. (1961). *Introduction to geometry*. New York, NY: Blaisdell.
- Enright, J. T. (1991). Paradoxical monocular stereopsis and perspective vergence. In S. R. Ellis (Ed.), *Pictorial communication in virtual and real environments* (pp. 567–576). London: Taylor & Francis.
- Gauss, C. F. (1827). Disquisitiones generales circa superficies curvas. Göttingen: Königlichen Gesellschaft der Wissenschaften. (Presented to the Göttingen Royal Society, October 6, 1827.) First published Commentationes Societatis Regiae Scientiarum Gottingensis Recentiores, vol. VI. (Reprinted in Carl Friedrich Gauss Werke, Volume 4, pp. 217–258).
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Hamm, J. (1983). *Drawing the head and figure*. New York, NY: Perigee.
- Hatton, R. G. (1910). *Figure drawing*. London: Chapman & Hall, Ltd.
- Hildebrand, A. (1893). *Das Problem der Form in der bildenden Kunst*. Strasbourg: Heitz.
- Hogarth, B. (1958). *Dynamic anatomy*. New York, NY: Watson-Guption.
- Horn, B. K. P., & Brooks, M. J. (1989). *Shape from shading*. Cambridge, MA: MIT Press.
- Koenderink, J. J. (1990). *Solid shape*. Cambridge, MA: MIT Press.
- Koenderink, J. J., & van Doorn, A. J. (1992). Surface shape and curvature scales. *Image and Vision Computing*, 10(8), 557–565.
- Koenderink, J. J., & van Doorn, A. J. (1995). Relief: pictorial and otherwise. *Image and Vision Computing*, 13, 321–334. [doi:10.1068/p130321](https://doi.org/10.1068/p130321)
- Koenderink, J. J., & van Doorn, A. J. (2003). Local structure of Gaussian texture. *IEICE Transactions on Information and Systems*, 86(7), 1165–1171. [doi:10.1023/A:1011126920638](https://doi.org/10.1023/A:1011126920638)
- Koenderink, J. J., & van Doorn, A. J. (2012). Gauge fields in pictorial space. *SIAM Journal on Imaging Sciences*, 5, 1213–1233. [doi:10.1137/120861151](https://doi.org/10.1137/120861151)
- Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1992). Surface perception in pictures. *Perception & Psychophysics*, 52, 487–496. [doi:10.3758/BF03206710](https://doi.org/10.3758/BF03206710)
- 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. [doi:10.1068/p3030](https://doi.org/10.1068/p3030)
- Koenderink, J. J., van Doorn, A. J., & Wagemans, J. (2011). Depth. *i-Perception*, 2, 541–562. [doi:10.1068/i0438aap](https://doi.org/10.1068/i0438aap)
- Lambert, J. H. (1760). *Photometria, sive, De mensura et gradibus luminis, colorum et umbrae*. Augsburg: V. E. Klett.
- Leonardo da Vinci (1452–1519). *Codex Urbinas*. Gathered together by Francesco Melzi before 1542. First printed in French and Italian as Trattato della pittura by Raffaello du Fresne in 1651.
- Lillholm, M., & Griffin, L. D. (2007). Statistics and category systems for the shape index descriptor of local image structure. *Image and Vision Computing*, 27(6), 771–781. [doi:10.1016/j.imavis.2008.08.003](https://doi.org/10.1016/j.imavis.2008.08.003)
- Lotze, H. (1852). *Medicinische psychologie oder physiologie der seele*. Leipzig: Weidmann'sche Buchhandlung.
- Mainardi, G. (1856). Su la teoria generale delle superficie. *Giornale dell'Istituto Lombardo*, 9, 385–404. [doi:10.1016/0315-0860\(79\)90075-2](https://doi.org/10.1016/0315-0860(79)90075-2)
- Metzger, W. (1975). *Gesetze des Sehens*. Frankfurt am Main: Verlag Waldemar Kramer.
- Pollack, P. (1955). A note on monocular depth-perception. *The American Journal of Psychology*, 68(2), 315–318. [doi:10.2307/1418907](https://doi.org/10.2307/1418907)
- Porteous, I. R. (1994). *Geometric differentiation for the intelligence of curves and surfaces*. Cambridge: Cambridge University Press.
- Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American*, 259, 76–83. [doi:10.1038/331163a0](https://doi.org/10.1038/331163a0)
- Rimmer, W. (1962, original 1877). *Art anatomy*. New York, NY: Dover.
- Rogers, L. R. (1969). *The appreciation of the arts. Vol. 2: Sculpture*. London: Oxford University Press.
- Sachs, H., (1990). *Isotrope Geometrie des Raumes*. Braunschweig: Friedrich Vieweg & Sohn.
- Schlosberg, H. (1941). Stereoscopic depth from single pictures. *The American Journal of Psychology*, 54(4), 601–605. [doi:10.2307/1417214](https://doi.org/10.2307/1417214)
- Schwartz, A. H. (1971). Stereoscopic perception with single pictures. *Optical Spectra*, 5, 25–27. [doi:10.1068/i0585](https://doi.org/10.1068/i0585)
- Smith, A. (2014). <http://www.assculpture.co.uk>. (last visited April 2014).

- Spelke, E. S. (2000). Core knowledge. *American Psychologist*, 55, 1233–1243. doi:10.1037/0003-066X.55.11.1233
- Spivak, M. (1999). *A comprehensive introduction to differential geometry* (Vol III, 3rd ed.). Houston: Publish or Perish.
- Strubecker, K. (1941). Differentialgeometrie des isotropen Raumes. I. Theorie der Raumkurven. *Akademie der Wissenschaften in Wien, Sitzungsberichte IIa*, 150, 1–53. doi:10.1016/j.cagd.2013.02.008
- Strubecker, K. (1942a). Differentialgeometrie des isotropen Raumes. II. Die Flächen konstanter Relativkrümmung $K = rt - s^2$. *Mathematische Zeitschrift*, 47, 743–777.
- Strubecker, K. (1942b). Differentialgeometrie des isotropen Raumes. III. Flächentheorie, *Mathematische Zeitschrift*, 48, 369–427. doi:10.1007/BF01180022
- Strubecker, K. (1944). Differentialgeometrie des isotropen Raumes. IV. Theorie der flächentreuen Abbildungen der Ebene. *Mathematische Zeitschrift*, 50, 1–92. doi:10.1007/BF01312437
- Thom, R. (1972). *Structural stability and morphogenesis*. San Francisco, CA: W. A. Benjamin.
- Vallortigara, G., Chiandetti, C., Rugani, R., Sovrano V. A., & Regolin, L. (2010). Animal cognition. *Wiley Interdisciplinary Reviews, Cognitive Science*, 1, 882–893. doi:10.1002/wcs.75
- van Doorn, A. J., Koenderink, J. J., Todd, J. T., & Wagemans, J. (2012). Awareness of the light-field: The case of deformation. *i-Perception*, 3, 467–480. doi:10.1068/i0504
- Yaglom, I. M. (1979). *A simple non-euclidean geometry and its physical basis*. New York, NY: Springer.



Jan Koenderink is a retired Professor of Physics. He is currently a guest Professor at the University of Leuven and the University of Utrecht. In the past he has worked on a variety of topics in physics, mathematics, computer science, biology, psychology, and philosophy. His post-retirement hobby is the relation (two-way, and both conceptual and historical) between the science of perception and the visual arts.



Andrea van Doorn is a retired Associate Professor. She is currently a guest Professor at the University of Leuven and the University of Utrecht. Her background is in physics, and she has a long standing interest in many topics of human perception, both empirically and theoretically.



Johan Wagemans has a BA in psychology and philosophy, an MSc and a PhD in experimental psychology, all from the University of Leuven, where he is currently a full professor. Current research interests are mainly in so-called mid-level vision (perceptual grouping, figure-ground organization, depth, and shape perception) but stretching out to low-level vision (contrast detection and discrimination) and high-level vision (object recognition and categorization), including applications in autism, arts, and sports (see www.gestaltrevision.be).