Relief Articulation Techniques

Jan Koenderink ¹,², Andrea van Doorn ², Liliana Albertazzi ³,⁴ and Johan Wagemans ¹,*

¹ Laboratory of Experimental Psychology, University of Leuven (KU Leuven), Belgium
² Experimental Psychology, Utrecht University, The Netherlands
³ Center for Mind/Brain Sciences (CIMeC), University of Trento, Italy
⁴ Department of Humanities, University of Trento, Italy

Received 25 April 2014; accepted 30 June 2014

Abstract
We consider techniques used in the articulation of pictorial relief. The related ‘cue’ best known to vision science is ‘shading’. It is discussed in terms of an inverse optics algorithm known as ‘shape from shading’. However, the familiar techniques of the visual arts count many alternative cues for the articulation of pictorial relief. From an art technical perspective these cues are well known. Although serving a similar purpose as shading proper, they allow a much flatter value scale, making it easier to retain the picture plane, or major tonal areas. Vision research has generally ignored such methods, possibly because they lack an obvious basis in ecological optics. We attempt to rate the power of various techniques on a common ‘shading scale’. We find that naive observers spontaneously use a variety of cues, and that several of these easily equal, or beat, conventional shading. This is of some conceptual interest to vision science, because shading has a generally acknowledged ecological basis, whereas the alternative methods lack this.

Keywords
Pictorial relief, shape, color, shape-from-shading

1. Introduction

‘Shape from shading’ has been one of the cues that was considered solved early in the development of computer vision (Horn and Brooks, 1989). The reason is that it relies on a very simple principle of radiometry known as Lambert’s law (Lambert, 1760). Lambert’s law states that the irradiance of a surface element varies with the cosine of the inclination of the surface with re-
spect to the direction of illumination. Granted a number of prior assumptions, one is left with a mathematical toy problem on which exists an extraordinary volume of literature. From an applications perspective, shape from shading is rarely useful, since the prior assumptions are unlikely to be met and it is next to impossible to check whether they are. Most tonal modulations one finds in a monochrome photograph of a natural scene can at best be described approximately by way of the canonical scheme, in many cases they call for a categorically different description.

In the classical visual art academies one taught ‘shading’ according to the ‘reception of the light’ (Clifton, 1973; Hogarth, 1981; Jacobs, 1988), a term due to Alberti (1435), in order to render surface relief. The portrait of a woman by Peter Paul Rubens (1577–1640), and the scene by Nicolas Poussin (1594–1665; Fig. 1) are generic examples.

In the study of visual perception one speaks of the ‘shading cue’. The bulk of studies of this cue has used a stimulus that has become conventional (Kleffner and Ramachandran, 1992; Ramachandran, 1988). It is a circular disk filled with a linear luminance gradient (Metzger, 1975). Such a disk tends to

Figure 1. At left Peter Paul Rubens, portrait of a woman. Rubens used classical shading according to the reception of the light. The illumination direction is from the upper left. Notice the resulting clear ‘reading’ of the surface relief. At right a wash drawing by Nicolas Poussin. This is still academic shading, with a well-defined direction of illumination (from the left), although there are no gradual transitions. Gradients are not required to construct strong pictorial relief, here Lambert’s law seems ‘binarized’, or ‘clipped’. Notice that this figure indeed captures the ‘shading’, but that many subtle color effects are missing in the reproduction. For instance, Rubens used a combination of red and black chalk, Poussin’s drawing is in wonderful sepia tone.
appear as the rendering of a sphere to most observers. In previous communications we have discussed the history of shape from shading studies in vision science (Erens et al., 1993; Wagemans et al., 2010). No doubt, the conventional stimulus arose in an attempt to isolate the shading cue, and present it in its purest form.

Although the conventional stimulus indeed ‘works’, in the sense that it evokes the awareness of a spherical relief in many observers (Metzger, 1975; Ramachandran, 1988), it is not at all the case that the shading cue may be regarded as understood. Indeed, the simplicity of the conventional stimulus is deceptive. It is also unlikely that the visual system somehow represents the radiometric processes, because one may easily come up with stimuli that fail to have an interpretation according to the radiometric model, yet evoke well-defined articulations of pictorial relief.

Examples of the latter abound in the visual arts, especially the arts from the eighteen-nineties till the nineteen-twenties. When artists dropped the concept of a painting as a ‘window’, but started to regard it as a planar assemblage of pigments, tonal painting soon became outmoded. Maurice Denis’ manifesto:

Remember that a picture, before being a battle horse, a nude, an anecdote or whatnot, is essentially a flat surface covered with colors assembled in a certain order. (Orig.: « Se rappeler qu’un tableau, avant d’être un cheval de bataille, une femme nue ou une quelconque anecdote, est essentiellement une surface plane recouverte de couleurs en un certain ordre assemblées. ») (Maurice Denis, 1890, p. 540)

of eighteen-ninety is a convenient anchor point. Of course, one still needed to suggest articulations of pictorial relief, but the means had to be adjusted so as to properly respect the picture plane. Various groups — and often generations — of painters sought to solve this in a variety of ways.

One way is to trade areas for edges. Edges are vital in painting, and essential in modulating relief. For instance here are some recommendations from a well-known teacher of oil painting:

There are two other kinds of edges — hard and soft. A hard edge is clear and distinct; a soft edge is fuzzy.

Edges are wonderful. . . by just varying your edges you can achieve a dimensional quality.

A soft edge shows continuity. . . . A hard edge shows ending.
(Cateura, 1995, p. 70)

It is possible to paint in terms of edges instead of areas. This is a great help in any attempt to retain the feeling for the picture plane, reason why edge-based
modulation became popular in the early nineteenth century. It is possible to think of a painting as a pure nexus of edges of various nature:

If edges are considered as field effects, rather than definite borderlines, they become zones of interchange. As such, they occur not only between verbally constituted entities, but everywhere. In the strictest sense, optical edges do not exist.
(Jacobs, 1988, p. 99)

Edge modulation is often combined with chromatic modulation, for a variety of reasons, on reason being the common need to retain a common value in some area in order to prevent it from — visually — ‘falling apart’. One substitutes chromatic modulations for value modulations:

One way to avoid making too many value changes as you add details is by making color temperature changes instead. That is, rather than make the highlight on a nose lighter so it stands out, make it cooler — and keep the cool value at the same value as the rest of the nose. Or, make the jawline recede-not by painting it darker, but by making it cooler. . . . Thus you preserve the original concept, the low-value look, of the portrait.
(Cateura, 1995, p. 18)

An ancient method is to evoke pictorial relief by edge darkening, where a narrow gradient of dark tone is used to suggest an occluding boundary. This technique has long been a favorite with sculptors, because it allows one to do away with incidental elements like the direction of illumination (Hogarth, 1981). This technique fully rejects references to radiometry (see Fig. 2). It can also be used in a ‘binary’, or ‘clipped’ fashion, then it looks like the edge modulations described by Pinna (1987). In many cultures, and stylistic periods, edge shading is preferred for, e.g., iconic images of saints, where the introduction of radiometric realism is often felt as somehow improper. In those images anything incidental would be inappropriate.

Edge shading easily suggests ‘bulk’, which is another reason why sculptors prefer this technique in their drawings. Therefore, in his book on shading, Burne Hogarth speaks of ‘sculptural light’:

Sculptural light is concerned primarily with three-dimensional form. Such light alludes to the sense of touch . . . Thus, sculptural light is often called . . . universal light . . . because it is an artificial creation . . .
(Hogarth, 1981, p. 82)

Indeed, edge darkening should not be confused with the shading obtained with frontal illumination. In frontal illumination the illumination direction and the
Figure 2. A drawing by Aristide Maillol (1861–1944) illustrates the ‘sculptural’ method of edge darkening. This has nothing to do with shape from shading proper. The light comes from nowhere, or is perhaps felt to derive from the artist’s and/or observer’s mind (as in Yeats (1936): *The mirror turned lamp*).

viewing direction approximately coincide. Taking a snapshot with the sun in the back is a common example. It yields flattish results, and is usually avoided by professional photographers for this reason (Adams, 1952; an exception is shown in Fig. 3). The effect is often seen in family pictures, taken with flash on the camera — a professional would separate the flash unit from the camera. Another familiar instance of frontal illumination is the full moon, which looks like a flat disk to most observers. Conventional shading does not at all approximate the effects of artistic edge darkening.

Instead of filling large areas with light or dark tone, one may largely, or even fully, limit the tonal modulation to the boundary between such areas (see online Supplementary Fig. S1). This works very well, and can easily be simulated in the laboratory. For want of a conventional term we refer to such contour modulation as bipolar edges. It is related to such familiar effects as the Cornsweet illusion (Cornsweet, 1970; Ratliff, 1965; Shapley and Gordon, 1985).

Yet another method substitutes chromatic variations for tonal ones. This became very common in the early twentieth century (see online Supplementary
Figure 3. This famous photograph by Edward Weston of 1934 clearly shows what Ansel Adams dubbed as the ‘limb effect’ (Adams uses this Weston example). Notice that the sun is right behind the photographer, the model covers her shadow on the ground. The edge darkening is mainly due to the non-Lambertian properties of her skin. The composition depends upon the similar values of the sand and her skin.

Fig. S2). We will speak of ‘chromatic modulation’ in contrast to ‘tonal modulation’. The modulations can be of various kinds. Chromatic modulation can almost do away with chiaroscuro (light/dark), and thus is a great help to the painter who desires to ‘respect the picture plane’. Such paintings are in relief, but are nevertheless ‘flat’. This possibility was much researched and exploited in the arts of the period straddling the year nineteen hundred.

Examples of such techniques abound, here we illustrate them with some paintings by Franz Marc (see online Supplementary Fig. S3). Chromatic modulation can be used to emulate proper shading, edge darkening, or contour modulation, as the case may be.

Vision science has little to say on the effectiveness of such ‘shading’ methods. Notice that ‘shading’ is really a misnomer here, as it might as well be for the conventional stimulus. One might more properly refer to them as a bunch of painting techniques used to achieve relief articulations. In this study we attempt a first — still somewhat limited — survey of the phenomenological facts, using a semi-quantitative method.

2. Rationale of the Study

The idea of the present study is to place the ‘articulation power’ of various, mutually very different, stimulus patterns in a common linear order. A simple order can be defined on the basis of the conventional stimulus by simply varying the contrast. As the contrast is diminished one becomes aware of a flattening of the relief. Phenomenologically it is a clear linear order, objectively it is parameterized by luminance contrast. This is very convenient and useful. The linear order used in the experiment is shown in Fig. 4. These images will
Figure 4. The case of proper shading. This also illustrates the nominal anchor points of the linear order pursued in the study. The scale is composed of the conventional shape from shading stimulus, with contrast varied in steps of a factor of two. Notice that the highest contrast looks almost spherical, whereas the lowest contrast yields an almost flat relief. Ideally, the powers of the other cues considered here should fall between the anchor points defined by these instances.

eventually become anchor points in the overall linear order. The idea is to use the shading cue as a ‘common currency’.

One group of stimuli of likely interest is a variation on the conventional shape from shading stimulus, with the tonal modulation replaced with chromatic modulation. In this study we use a green–red modulation about yellow, and a yellow–blue modulation about gray, as substitutes for a white–black modulation about gray (see online Supplementary Figs S4 and S5).

Of course, the definition of ‘chromatic modulation’ is of importance here. In vision research one would immediately consider equi-luminant patterns. However, from the perspective of the visual arts this makes little sense. For instance, consider a yellow–blue contrast. In the equiluminant condition the ‘yellow’ would necessarily have to be downgraded to a dark brown, in order to match the blue in luminance. The problem is caused because paintings necessarily display colors in context, so light and dark play complementary roles (Fig. 5). Pure ‘aperture colors’ play no role at all. Consequently, there is hardly a use for the concept of luminance.

Consequently, we did not use ‘equiluminant’ in the CIE definition (Koenderink, 2012; Stokes et al., 1996), which sets the red–green–blue ratios roughly at 3:6:1, because it addresses an essentially color-blind visual subsystem that mainly caters for movement. Indeed, equiluminance is usually established by way of flicker photometry, or ‘motion nulling’ (Gregory, 1985; Kaiser et al., 1989; Webster and Mollon, 1993). For the present purpose, we rather need to balance hues with respect to their Gestalt grouping power (see online Supplementary Fig. S6). This requires red–green–blue ratios close to 1:1:1. It perhaps reminds one of Schopenhauer’s (1816) ‘parts of daylight’, where red, green, and blue are treated as ‘equal parts’. The cardinal colors red, green, blue, cyan (that is blue and green), purple (that is blue and red), and yellow (that is red and green) indeed have very similar weights in Gestalt groupings. Stimuli produced in this way even at first sight look much closer to
Figure 5. In this Yin–Yang pattern the white and black areas play equivalent roles. Evidently, ‘equiluminance’ makes no sense here! In terms of pictorial composition white and black on a medium gray ground carry the same ‘weight’. This is generally true in cases where two colors — like white and black here — play together, or compete with each other, in Gestalt configurations. The colors are not ‘equiluminant’, but of ‘equal weight’.

the chromatic modulations as used by painters than equiluminant renderings do.

The basic ‘radiometric’ structure may be dropped altogether. One way that is known to work in monochrome is based on contour modulation by way of bipolar edges (Georgeson et al., 2007; Hesse and Georgeson, 2005; Sun and Schofield, 2012 — see Fig. 6). Such a stimulus can easily be emulated in terms of chromatic instead of tonal modulation (Ejima and Takahashi, 1988; Kingdom, 2003; Livingstone and Hubel, 1987 — see online Supplementary Fig. S7).

The method of edge darkening is illustrated in Fig. 7. Notice that such shading — as it is frequently called, which has no radiometric origin at all — is quite effective in suggesting ‘volume’. This stimulus is also simply rendered chromatically instead of tonally (see online Supplementary Fig. S8).

We also introduce a number of stimuli that will in all probability be perceived as flat (Fig. 8). They are needed to establish a rock bottom for the linear order. We use both tonal and chromatic modulations (see online Supplementary Fig. S9). Such patterns are immediately seen as flat, although they may be perceived as recessions or as pedestals. Indeed, that is how they are often used in the visual arts. You can ‘color’ an area by simply modulating its boundary. In the nineteen-sixties, this was interpreted by Ratliff (1965) in terms of lateral inhibition. Ratliff (1965) gives a variety of instances in the arts of various cultures.

The effects of chromatic modulations in Gestalt-like configurations have not been studied into great detail (Ejima and Takahashi, 1988; Takahashi et
Figure 6. Example of contour modulation by way of bipolar edges. The linear gradient inside the disk has been replaced with a uniform tone. The tonal area modulation (‘shading proper’) has been limited to a narrow strip about the circular outline. This clearly articulates pictorial relief.

Figure 7. The case of edge darkening. The linear gradient inside the disk has been replaced with a tonal edge darkening. This sculptural shading — as it is often called — has no radiometric origin, yet it is quite effective. It is common in many cultures, and stylistic periods, because it does not introduce an irrelevant accident like a direction of illumination.

al., 2010; de Weert and Spillmann, 1995), despite their importance in the visual arts. A common wisdom in vision science is that ‘equiluminant’ chromatic modulations prevent depth and shape cues to be effective (Livingstone and Hubel, 1987). However, other studies have not invariably found this (Cavanagh, 2009; Cavanagh et al., 1992; van Doorn et al., 2005). This is at least partly due to the operational definition of equiluminance. Moreover, as argued above, the concept of equiluminance can hardly be applied to paintings in a sensible manner. What really counts are the powers in Gestalt formation of value and chroma.
Figure 8. Two tonal stimuli that will almost certainly be perceived as ‘flat’, although possibly raised or recessed from the picture plane.

It should not be concluded that the set of stimuli introduced here exhausts the tools used by visual artists that allow them to modulate pictorial relief. One simple and important tool omitted here is vignetting (Koenderink and van Doorn, 1983, 2004; Langer and Bülthoff, 2000). Vignetting is a radiometric effect that differs from shading, and relies on the fact that deep recesses will appear dark. Vignetting has also chromatic implications, for instance, concavities on a Caucasian nude will become warmer. Thus a sculptor may ‘paint’ a pupil in marble with a drill, the painter uses dark blobs to recess eye sockets, and so forth (Sun and Schofield, 2012). There are still other techniques in regular use. The topic is one that invites attention in vision science. A blurry dark blob will immediately look like a recession, often used in painting to indicate the eye sockets in distant faces. On the other hand, a blurry light blob tends to appear as a bulge.

The images shown in Figs 4, 6, 7, 8 and S4, S7, S8, S9 comprise the set used in the experiment. In the text we use the descriptions of Table 1. Stimuli are compared pairwise. The idea is to judge from each pair which image evokes the strongest articulated relief. From such judgments one may attempt to find a linear order. Since the images that make up the linear order are located in such an order (Fig. 4), the order becomes ‘calibrated’ by them. The scale items function as anchor points, allowing an ordinal comparison of the various items.

Since observers have to make many more judgments than there are items in the linear order, such a procedure also yields an indication as to whether a description by way of a linear order makes sense. In case it does not, the best linear order will fail to explain the actual judgments, so we have an immediate statistical check. This is important, since most of the stimuli do not really have an obvious relation to ‘shading’ at all.
Table 1.
The stimuli used in the experiment

<table>
<thead>
<tr>
<th>Index</th>
<th>Stimulus structure</th>
<th>Illustrated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Shading MON I</td>
<td>Figure 4</td>
</tr>
<tr>
<td>2</td>
<td>Shading MON II</td>
<td>Figure 4</td>
</tr>
<tr>
<td>3</td>
<td>Shading MON III</td>
<td>Figure 4</td>
</tr>
<tr>
<td>4</td>
<td>Shading MON IV</td>
<td>Figure 4</td>
</tr>
<tr>
<td>5</td>
<td>Shading R–G</td>
<td>Figure S4</td>
</tr>
<tr>
<td>6</td>
<td>Shading Y–B</td>
<td>Figure S4</td>
</tr>
<tr>
<td>7</td>
<td>Edge darkening MON</td>
<td>Figure 7</td>
</tr>
<tr>
<td>8</td>
<td>Edge darkening R–G</td>
<td>Figure S8</td>
</tr>
<tr>
<td>9</td>
<td>Edge darkening Y–B</td>
<td>Figure S8</td>
</tr>
<tr>
<td>10</td>
<td>Bipolar edge MON</td>
<td>Figure 6</td>
</tr>
<tr>
<td>11</td>
<td>Bipolar edge R–G</td>
<td>Figure S7</td>
</tr>
<tr>
<td>12</td>
<td>Bipolar edge Y–B</td>
<td>Figure S7</td>
</tr>
<tr>
<td>13</td>
<td>Flat light MON</td>
<td>Figure 8</td>
</tr>
<tr>
<td>14</td>
<td>Flat light R–G</td>
<td>Figure S9</td>
</tr>
<tr>
<td>15</td>
<td>Flat light Y–B</td>
<td>Figure S9</td>
</tr>
<tr>
<td>16</td>
<td>Flat dark MON</td>
<td>Figure 8</td>
</tr>
<tr>
<td>17</td>
<td>Flat dark R–G</td>
<td>Figure S9</td>
</tr>
<tr>
<td>18</td>
<td>Flat dark Y–B</td>
<td>Figure S9</td>
</tr>
</tbody>
</table>

3. Methods

3.1. Participants

The 15 participants were students from Trento University who participated in order to gain credits for their study program. They were 14 female, one male, in their early twenties. All had normal, or corrected to normal vision, and had no color deficiencies. They were naive with respect to the experiment, and had no experience in vision research, nor had they formal training in painting or drawing techniques. All participants had Italian as their first language. The experiment was conducted after obtaining informed consent from the participants.

3.2. Experimental Procedure

The stimuli were presented on the LCD screen of an Apple Powerbook. The screen was previously calibrated via the SuperCal program. We used a linearized (gamma 1.0) color table. The color temperature of the white was 5700°K. The screen measures 1400 × 900 pixels. The screen background pattern had about 2000 polygonal cells of random shape and gray values. The observer sat at 57 cm in front of the screen in an otherwise darkened room. The response device was a keyboard.
3.3. Stimulus Configuration

Each stimulus pattern was $512 \times 512$ pixels square. They were presented juxtaposed, the pair being centered on the screen.

3.4. Task

The observer had to indicate whether the left or the right image gave rise to a more articulate pictorial relief by hitting the appropriate arrow key. In case both images appeared flat the space bar was used.

3.5. Experimental Data

We obtained order judgments for all pairs. This enables the fit of a best linear order. The goodness of fit provides a useful measure of consistency.

4. Results

In the experiment we presented all pairs, in random order. The participant had to decide which of the two images had the most articulated pictorial relief. The participant used the arrow keys of the keyboard to indicate ‘left’ or ‘right’.

There are eighteen distinct instances, of which four are flat. This yields 324 pairs, since all ordered pairs are presented, including ones in which the two items are equal. Of these cases where both of the stimuli are flat there are sixteen presentations, thus about one in twenty presentations shows two flats. Observers always immediately identify these cases. We included a category ‘both flat’, to be indicated by pressing the space bar.

Observers find the task easy, and even fun to do. On the average they take about three-quarters of a second for a single response. They do all 324 responses in a single session.

We included the possibility to respond ‘both flat’, which is certainly unusual. We did this, because we really need the ‘flat’ judgment as base line. We have no need for observers to thoughtlessly generate random responses to stimuli that do not evoke any feelings of pictorial shape. On the contrary, we want to be notified of that fact. These trials may be regarded as ‘catch trials’. If only one of a pair of stimuli is obviously flat, we expect observers to respond that the other one has greater three-dimensional articulation. Thus we obtain an estimate of mere ‘sloppiness’ of observers. In the final analysis the catch trials are ignored.

Since we have a judgment on any pair, it is an easy matter to find the best fitting linear order for each participant. One simply counts in how many cases a given stimulus was judged as more articulate than the accompanying mate. This yields a number for each stimulus; this number is used to determine the order. It is essentially a ‘voting’ procedure. Perhaps surprisingly, this algorithm can be shown to be equivalent to a certain least squares method (van
From such an order one may retrodict the expected judgment for any pair. This again, can be compared with the actual judgments, thus yielding a useful *figure of merit* as the difference between the number of concordant and the number of discordant pairs, divided by the total number. Similar to Kendall’s tau, this yields a number between minus and plus one. A value of one implies perfect consistency, whereas random responses will yield zero on the average. We find a median value of 0.47, the interquartile range being (0.39, 0.54).

From simulation studies we know that the standard deviation of the figure of merit for a random observer is about 0.08, thus the responses are highly significant. We define a ‘random observer’ as one who is able to identify the really flat patterns as such, but decides upon the most articulated member of a pair of generic ones with equal probability. Simulation is fast, and one readily samples a hundred thousand trials, yielding well-defined statistics.

The lowest figure of merit is 0.24, which is still highly significant. However, the observers in the below 0.25 quartile range are evidently responding inconsistently to some degree. The maximum value for the figure of merit is 0.64. We discuss the observers in the low quartile in more detail below.

The voting orders of the individual observers mutually correlate with Kendall’s tau in the range (0.19, 0.88), the median is 0.57, interquartile range (0.46, 0.65). These rank correlations can be compared with the individual figures of merit of the observers. We find that the rank correlation with the least consistent member of the pair is 0.2.

The voting orders for the individual observers correlate (Kendall’s tau) well with the median voting order (Fig. 9). The range is (0.53, 0.83), median 0.66,

![Figure 9. Scatter plot of all individual voting orders against the median of the voting orders.](image)
interquartile range (0.62, 0.76). Moreover, the rank order correlations also correlate with the order of figure of merits (Kendall’s tau 0.26). Thus the more consistent an individual observer, the better that observer correlates with the overall order. This indicates that the overall order represents the opinio communis of the individually most consistent participants.

Given this order for each observer we proceed as follows. We find the order of the first objectively flat stimulus, and curtail the order at that. Stimuli judged less articulate than any flat one are thus simply discarded as ‘flat’. This is a very harsh criterion, since a single slip might render an observer less discriminative, which has a strong effect on the result. We prefer such a criterion because we were working with a group of naive observers. A harsh criterion may be regarded as an objective way to get rid of inevitable outliers. The most discriminative observer ranks eleven of the stimuli as articulate, whereas the least discriminative observers rank as few as three stimuli as articulate. The median number is six, with interquartile range (4.0, 8.8). From simulation results over ten thousand simulated sessions we know that a random observer scores a median of one, interquartile range zero to two. Thus the responses are highly dependable. These results are shown in Table 2 and Fig. S10.

In order to construct an overall order, we employ a voting procedure. This involves nothing more than counting. The most articulate stimulus is the one

<table>
<thead>
<tr>
<th>Table 2.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The individual results (ranked stimuli), ordered by discriminative power of the observers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7, 1, 10, 9, 2, 6, 12, 8, 11, 3, 5</td>
</tr>
<tr>
<td>2</td>
<td>1, 2, 6, 3, 12, 10, 11, 7, 5</td>
</tr>
<tr>
<td>3</td>
<td>11, 1, 10, 2, 12, 6, 3, 5, 7</td>
</tr>
<tr>
<td>4</td>
<td>7, 10, 1, 9, 12, 11, 2, 8, 6</td>
</tr>
<tr>
<td>5</td>
<td>6, 1, 12, 11, 10, 2, 5, 7</td>
</tr>
<tr>
<td>6</td>
<td>1, 7, 12, 11, 10, 9, 2, 6</td>
</tr>
<tr>
<td>7</td>
<td>7, 1, 12, 2, 11, 10, 3, 9</td>
</tr>
<tr>
<td>8</td>
<td>1, 10, 12, 11, 2, 7</td>
</tr>
<tr>
<td>9</td>
<td>7, 1, 10, 2, 12, 11</td>
</tr>
<tr>
<td>10</td>
<td>10, 12, 11, 7</td>
</tr>
<tr>
<td>11</td>
<td>10, 7, 1, 12</td>
</tr>
<tr>
<td>12</td>
<td>10, 7, 11, 12</td>
</tr>
<tr>
<td>13</td>
<td>7, 9, 1, 12</td>
</tr>
<tr>
<td>14</td>
<td>7, 1, 9</td>
</tr>
<tr>
<td>15</td>
<td>7, 10, 1</td>
</tr>
</tbody>
</table>

Note. The numbers refer to the indexed stimuli listed in Table 1.
Table 3.
The final result (ranked stimuli) obtained by majority vote. The ‘anchor stimuli’ have been emphasized by boldface

<table>
<thead>
<tr>
<th>Rank order #</th>
<th>Stimulus structure</th>
<th>Stimulus index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Edge darkening MON</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td><strong>Shading MON I</strong></td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Bipolar edge MON</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Bipolar edge R–G</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>Shading Y–B</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Bipolar edge Y–B</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>Edge darkening Y–B</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td><strong>Shading MON II</strong></td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td><strong>Shading MON III</strong></td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>Shading R–G</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>Edge darkening R–G</td>
<td>8</td>
</tr>
</tbody>
</table>

Note. The indexed stimuli are listed in Table 1.

that is most often found at the top of the individual orders, and so forth. This yields the order illustrated in Table 3 and Fig. S11. We end up with an order of eleven instances, the remaining instances being considered ‘flat’. The flats include the lowest contrast shaded stimulus, thus our basic scale turns out to be long enough.

The tonal edge darkening image heads the list, immediately after it is the conventional shading stimulus of highest contrast. The tonal contour modulation stimulus occurs before the next anchor in the scale, so do four of the chromatic stimuli. These are apparently excellent replacements for the standard shape from shading technique to evoke relief articulation. The chromatic stimuli score quite high, especially the yellow–blue variety.

We determined inter-observer rank correlations. On the whole the rank correlations — that is the Kendall tau for the set intersection — are satisfactory, with a single exception. A case with negative rank correlation is surely spurious, and we would consider ignoring this observer as ‘outlier’. We know from previous studies that a fair fraction of the general population fails to use the ‘shading cue’ to some degree, or even totally (van Doorn et al., 2012). Such observers report that they experience the conventional shading stimulus as a flat gradient of tone. We also met with reviewers of earlier manuscripts who were of the same opinion. Perhaps surprisingly, this seems not to be recorded in the literature. Thus ignoring a fraction of outliers certainly makes sense.

We found that omitting up to one-quarter outliers by various reasonable criteria it did not affect the overall order. Apparently the estimate is quite robust; hence our decision to simply retain all data.
Given the overall order (Table 3), we may compare the individual orders to it (Table 2). The range of the resulting rank correlations is \((-0.33, 1.00)\), the median 0.33, interquartile range (0.22, 0.59), which is acceptable. Of course, throwing out a fraction of outliers will raise these numbers very markedly, which would perhaps offer a more realistic perspective on the data.

5. Discussion

Pictorial relief is a visual quality. That is to say, we are dealing with first person reports. Thus it cannot be studied by way of the familiar methods of psychophysics, or ‘dry physiology’, but only as ‘experimental phenomenology’ (Koenderink, in press). In comparing a chromatic contour modulated rendering to a true shaded rendering of a convex relief (‘sphere’), there is no way such stimuli could ever be said to be ‘the same’, for any contrast of the shaded one. Thus the case is similar to, for instance, the equibrightness judgment, in which case the colors to be compared have different hues. Such colors will never look the same, thus the judgment of equibrightness is technically a ‘first person report’.

Indeed, for the hard-core scientist such concepts as equibrightness are nonentities. Of course, in the same mind-frame ‘art’ is a nonentity. Thus one should accept that virtually all studies of ‘art and perception’ of conceptual interest are in the realm of experimental phenomenology. This certainly applies to our study.

Since visual qualities are essentially subjective, there is no way to ascertain objectively what is in the participant’s visual awareness. Of course, three of the authors — none of them used in the group of naive observers — performed the task themselves, so we have their first person reports — or at least each of these authors has their own. Their orders are almost identical to the outcome of the experiment, and all three had figures of merit that would put them into the top quartile of the group of participants. The latter fact may have to do with their long experience in ‘mindful looking’ and motivation, whereas it is likely that at least some of the participants let their minds wander while performing the task. None of the participants had ever learned to look mindfully, as in a formal painting, or drawing training. Thus we are interested in ‘grading’ our group of participants, both with respect to internal self-consistency, and with respect to inter-subject consistency.

With respect to inter-subject differences, we know from previous experiments that a fair fraction of the population has difficulties using the shading cue, with a small fraction that even does not seem to use it at all (van Doorn et al., 2012). There appear to be varieties here, such as the specific inability to see concavities. Such variations in the normal population have never been fully mapped out. Thus it is \textit{a priori} not unlikely that there will be subgroups
of participants with mutually distinct visual awarenesses. We also desire to ‘grade’ participants with respect to that.

Both types of grading can in principle been done in an objective way. What is not possible is to identify the nature of the visual awareness — the visual qualities and meanings — of such subgroups. Of course, a high degree of concordance with the results of an author who did the experiment perhaps suggests that the latter’s first person report might apply to that of a subgroup. All one can do here is try to maximize the acceptability of such an interpretation. Interpretations are necessarily subjective, whereas the data are not less objective than those typically considered in classical psychophysics.

The internal self-consistency of the participants varies widely. This may be due to lapses of concentration, or just sloppiness, or it might have a cause in a different microgenesis of pictorial relief. A rank correlation of the voting orders of pairs of observers also shows a large spread. This will partly be due to the range of self-consistencies, but there might be a contribution of actual differences in visual awareness. Finally, the ‘resolution’ of the participants ranges from rather shallow to quite high: the most discriminative observer resolving eleven steps before merging into the flat category, the least discriminative as few as three. Again, possible explanations range from degrees of sloppiness to differences in visual awareness.

Table 4 shows an overview of the individual figure of merit (FOM), resolution (number of levels of articulation above flat ‘N3D’), and rank correlation (Kendall’s tau) of the observer with the overall order, for all fifteen participants.

We also checked ‘sloppiness’ as such. It turns out to be the case that most ‘sloppy’ observers also have low figures of merit, low resolution, and marginal correlation with the mean. Thus there is much reason to think that they had problems with ‘looking mindfully’. Of course, this is only to be expected with persons who never had formal training in looking. It is reassuring that they still conform to the majority, although marginally. Similar considerations probably apply to the bulk of the crowd in art museums, who may spend up to a few seconds to a painting (Elkins, 2013).

This table thus includes the participants with low internal self-consistency, those with low correlation with the overall result, as well as those with only low resolution. There are some observers with reasonable internal self-consistency, and reasonable resolution that nevertheless correlate hardly with the overall order. Thus there exists some basis for the notion that there might be real differences in awareness. The possible nature of such differences is perhaps best studied in the individual orders shown in Table 2.

Given the limited size of the group we refrain from such interpretations here, and focus on the overall picture. The overall result is quite clear. True shading is not the only way to articulate pictorial relief. ‘Edge darkening’,
Table 4.
An overview of the individual figure of merit (FOM), resolution (number of levels of articulation above flat ‘N3D’), and rank correlation (Kendall’s tau) of the observer with the overall order

<table>
<thead>
<tr>
<th>FOM</th>
<th>Levels</th>
<th>Kendall tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.64</td>
<td>11</td>
<td>+0.60</td>
</tr>
<tr>
<td>0.59</td>
<td>6</td>
<td>+0.60</td>
</tr>
<tr>
<td>0.58</td>
<td>4</td>
<td>+0.33</td>
</tr>
<tr>
<td>0.55</td>
<td>4</td>
<td>+0.33</td>
</tr>
<tr>
<td>0.53</td>
<td>9</td>
<td>+0.56</td>
</tr>
<tr>
<td>0.53</td>
<td>8</td>
<td>+0.50</td>
</tr>
<tr>
<td>0.48</td>
<td>3</td>
<td>+1.00</td>
</tr>
<tr>
<td>0.47</td>
<td>6</td>
<td>+0.20</td>
</tr>
<tr>
<td>0.47</td>
<td>4</td>
<td>+0.67</td>
</tr>
<tr>
<td>0.45</td>
<td>8</td>
<td>+0.50</td>
</tr>
<tr>
<td>0.44</td>
<td>9</td>
<td>+0.00</td>
</tr>
<tr>
<td>0.38</td>
<td>8</td>
<td>+0.07</td>
</tr>
<tr>
<td>0.27</td>
<td>9</td>
<td>+0.28</td>
</tr>
<tr>
<td>0.26</td>
<td>4</td>
<td>−0.33</td>
</tr>
<tr>
<td>0.24</td>
<td>3</td>
<td>+0.33</td>
</tr>
</tbody>
</table>

Note. This compilation suggests various interesting possibilities. An extensive population study might well be rewarding.

contour modulation via ‘bipolar edges’, and several of the chromatically modulated variations are just as effective, or even more so. Of course, there are minor variations between observers. Only three out of fifteen observers have the true shading heading their list.

6. Conclusions

The general conclusion is that there is nothing special about the true shading cue, the one purportedly exploiting ecological physics in terms of Lambert’s law. Several of the methods in actual use with painters have hardly any, or at least a different, relation to the ecological physics of radiance distributions, yet yield equally strong, or even stronger cues to pictorial relief articulation.

Perhaps surprisingly, the pre-renaissance ‘edge darkening’ technique actually beats Alberti’s ‘reception of the light’ in its efficaciousness of evoking the awareness of articulate relief. Indeed, photographic realism apparently has little to do with the techniques to evoke pictorial relief. Evidently non-realistic, but conventional methods like contour modulation, or the substitution of chromatic for tonal variation, are very effective.

There are hardly ecological explanations for the efficaciousness of evidently non-realistic methods. It is also unlikely that our participants called on their
familiarity with cultural conventions. The only case where the latter might apply is perhaps the edge darkening. This is a technique that has ancient roots in the Western cultural heritage to which our observers belong. Edge darkening has been used throughout the ages, and indeed to the present time. It is often preferred in varieties of religious art, because it does not depend upon such arbitrary incidents as illumination direction. This is exactly the same reason why this technique is so often preferred by sculptors for their own use.

In view of this, one may well wonder whether the standard textbook explanations of the shading cue capture the essence of the phenomenon. If so many different techniques work just as well, or even better, then what makes the ‘correct’ shading cue so special? It may well be the case that the whole notion of shape from shading is spurious (Koenderink et al., 2013), and that biological vision research should leave it to the computer vision engineers. From a historical point of view, this is a topic where the visual arts have gone far beyond scientific investigation. Especially the tumultuous developments in the visual arts from the eighteen-nineties to the nineteen-thirties may be mined as a major exploration of the phenomenology of vision. It is somewhat of an embarrassment that vision research so far has missed a large set of remarkably potent ‘relief cues’. We suggest that the very term ‘shading cue’ should be dropped for something like ‘cues for relief articulation’.

If this is accepted, then the conventional ‘explanation’ of the shading cue in terms of ecological physics is put in jeopardy. If it is indeed acknowledged that the other cues do not immediately admit of such an ecological basis, then perhaps all of them — including the shading cue — are more properly understood as templates of the human optical user interface.

This is exactly what the painter is exploiting. The painter arranges colors on a planar surface such as to evoke certain visual experiences in observers. This can be done in infinitely different ways. Some — like ‘shape from shading’ — may be immediately traced to ecological optics, whereas for others the relation to biological fitness may appear remote. But, of course, this does in no way stop the painter from using anything that effectively triggers the generic optical user interfaces of fellow humans.

Acknowledgements

The work was conducted at CIMeC (University of Trento), October 2013, during the visiting project of Jan Koenderink. The research was also supported by the Methusalem program by the Flemish Government (METH/08/02), awarded to Johan Wagemans. Carmen Granito assisted in running the experiments.
References

   A large-scale technique and observations, Perception 21, 219–226.
Cavanagh, P. (2009). All’s jazzy and unstable on the colour front: the impact of Gregory’s
   pioneering paper on vision at equi-luminance, Perception 38, 831–832.
Ejima, Y. and Takahashi, S. (1988). Illusory contours induced by isoluminant chromatic pat-
   terns, Vis. Res. 28, 1367–1377.
   shading, Percept. Psychophys. 54, 145–156.
   features: scale–space analysis of edge and blur coding in human vision, J. Vis. 7, 7, 1–21.
Gregory, R. L. (1985). Movement nulling: for heterochromatic photometry and isolating chan-
   nels for ‘real’ and ‘apparent’ motion, Perception 14, 193–196.
   191.
   Psychophys. 52, 18–36.
Koenderink, J. (in press). Methodological background: experimental phenomenology, in: Ox-
   ford Handbook of Perceptual Organization, J. Wagemans (Ed.). Oxford University Press,
   Oxford, UK.
Koenderink, J. J. and van Doorn, A. J. (1983). Geometrical modes as a general method to treat
Koenderink, J. J. and van Doorn, A. J. (2004). Shape and shading, in: The Visual Neurosciences,
   299–302.


Sun, P. and Schofield, A. J. (2012). Two operational modes in the perception of shape from shading revealed by the effects of edge information in slant settings, *J. Vis.* 12, 1–21. DOI: 10.1167/12.1.12.


