Light fields and shape from shading

Andrea J. van Doorn
Faculty of Industrial Design, Delft University of Technology, Delft, The Netherlands

Jan J. Koenderink
EEMCS and The Flemish Academic Centre for Science and the Arts, Delft University of Technology, Delft, The Netherlands, & Laboratory of Experimental Psychology, University of Leuven (K.U. Leuven), Leuven, Belgium

Johan Wagemans
Laboratory of Experimental Psychology, University of Leuven (K.U. Leuven), Leuven, Belgium

From a theoretical point of view, the use of the shading cue involves estimates of the light field and thus observers need to judge the light field and the shape simultaneously. The conventional stimulus in perceptual experiments, a circular disk filled with a monotonic gradient on a uniform surround, represents a local shading or tonal gradient. In typical scenes, such gradients vary smoothly from point to point over large areas, whereas light fields are globally defined and tend to be invariant over large parts of the scene. Hence, it is hardly surprising that multi-local shape estimates tend to synchronize although previous reports of such synchronies involved uniform, homogeneous light fields. Here, we consider more complicated and more realistic light fields. We present extensive, highly structured, quantitative observations using novel paradigms. Human observers are able to deal with some structured light fields but totally fail in others, even though these may be formally similar (like radial and circular fields). Observers respond very differently in some cases where the light fields differ only by sign, like converging and diverging fields. These results can be qualitatively understood on the basis of a few simple assumptions, mainly global top-down template matching of peripheral local data.

Keywords: light fields, illuminance flow, shape from shading, perceptual ambiguity, switching, synchronization, template matching

Citation: van Doorn, A. J., Koenderink, J. J., & Wagemans, J. (2011). Light fields and shape from shading. *Journal of Vision, 11*(3):21, 1–21, http://www.journalofvision.org/content/11/3/21, doi:10.1167/11.3.21.

Introduction

"Shape from shading" is a term that derives from algorithmic or machine vision (Forsyth & Ponce, 2002; Horn & Brooks, 1989; van Diggelen, 1959; Zhang, Tsai, Creyer, & Shah, 1999). It refers to a class of algorithms that computes three-dimensional shape on the basis of image structure. Of course, this involves numerous assumptions, some generic, some limiting, some incoherent (i.e., in conflict with physics). Typical assumptions include uniform, homogeneous light fields, Lambertian surfaces of uniform albedo, and the absence of both vignetting and multiple scattering. Light fields are often confined to a collimated beam, such as direct sunlight. Lambertian surfaces (Lambert, 1760) do not exist in the strict sense, though white plaster or paper comes close. Vignetting (Forsyth & Zisserman, 1991; Koenderink & van Doorn, 1983, 2003b) is the result of occlusion of the source by the scene itself. Cast and body shadows are the simplest examples. More complicated, and often important, cases involve extended sources and non-convex objects. Similar constraints apply to multiple scattering, which depends upon the possibility of disjunct surface elements to "see each other." In order to get rid of vignetting and multiple scattering, one would have to limit the scene to a single convex object—say an egg in outer space—and consider only the illuminated side. Of course, this is rarely of much interest. More realistic instances include shallow reliefs; in such cases, both vignetting and multiple scattering become negligible. Many scenes contain areas of this type because smoothly curved surfaces approximate flattish relief when considered in sufficiently small areas. Given these assumptions, the luminance sampled by the camera will reflect the illuminance of the surfaces in the scene (though the factor of proportionality is undefined), and one may invoke Lambert's cosine law (Lambert, 1760) to make the connection with scene geometry. The angle at which the beam strikes a surface determines its illuminance, thus the sampled luminance variations reflect surface attitude variations with respect to the direction of the illuminating beam. Although the shading is insufficient to specify scene geometry fully in this way, one at least finds solutions *modulo* a group of ambiguities (Belhumeur, Kriegman, & Yuille, 1999). These ambiguities include absolute distance,

doi: 10.1167/11.3.21 Received July 29, 2010; published March 29, 2011 ISSN 1534-7362

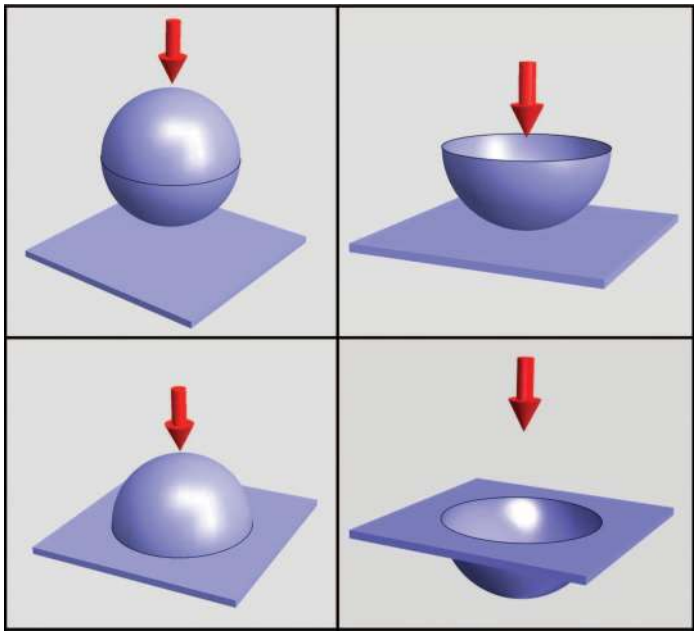

Figure 1. An inventory of edges and outlines. The viewing direction is along the red arrow; thus, one sees an object with circular boundary against a uniform background in all cases. However, the nature of the boundary can vary a lot. From left to right, top to bottom, one has an "occluding contour" (sphere in front of a plane), a "surface boundary" (cup in front of a plane), a "dihedral edge" (a protrusion rising from a plane, the "cap"), and a "dihedral edge" (a depression in a plane; the "cup").

overall spatial attitude ("additive plane"), and depth of relief ("bas-relief ambiguity"; Belhumeur et al., 1999). Such ambiguities also occur in human perception (Brewster, 1832; Hill & Bruce, 1994; Kleffner & Ramachandran, 1992; Ramachandran, 1988a, 1988b; Rittenhouse, 1786).

In the case of human vision, one speaks of the "shading cue" (Palmer, 1999). The first (proto-) scientific treatments originate from the visual arts (Baxandall, 2005). The eighteenth and nineteenth art academies invariably had a "cast room" with white plaster casts of antique sculpture. Students were required to spend several years sketching on a daily basis in order to master the art of *chiaroscuro* (or *clair-obscur*), that is, the art of shading. They would also receive some verbal instruction and lectures from some master to get them on their way. Eventually, they would graduate to the life drawing classes. Nude models are less convenient study objects than casts because they move (at least somewhat) and have non-Lambertian surfaces. (Most real surfaces are non-Lambertian; Dana, van Ginneken, Nayar, & Koenderink, 1999.) The posing stage would be illuminated in various standard, well-designed manners, a bit like modern Hollywood studios. The "shading cue" was indispensable because of a certain way to add "relief" to otherwise flattish (cartoon) drawings. Literature on the "reception of the light" starts with Alberti's "De Pictura" (http://www.noteaccess.com/Texts/Alberti) and Leonardo's notebooks (http://www.gutenberg.org/etext/5000). The truly scientific literature dates from the early twentieth century and is mainly in the phenomenological, Gestalt tradition (Luckiesh, 1916; Metzger, 1975; Turhan, 1935). Modern work (Palmer, 1999; Ramachandran, 1988a, 1988b) has adopted this. Perhaps unfortunately, attempts to use shape-from-shading algorithms as models for human visual abilities have proved unsuccessful.

The conventional stimulus in experimental psychology of human perception of shape from shading has been the linear gradient, contained in a circular disk (Figure 2, left). Such a pattern represents shading in its utmost abstraction, a mere local, linear gradient; it is a well-considered, though extreme, form of stimulus reduction.[1] The one feature that detracts from its success is the sharp circular outline that has nothing to do with shading but represents a cue in its own right (Cate & Behrmann, 2010; Hayakawa, Nishida, Wada, & Kawato, 1994; Humphrey, Symons, Herbert, & Goodale, 1996, Norman & Raines, 2002). Depending on whether it is seen as an occluding contour, a dihedral edge, or a surface boundary, one is aware of a sphere (necessarily convex), a local depression ("cup") or protrusion ("cap") in a plane, or a hemispherical cup with free surface boundary (Figure 1). One way to remove this confounding cue is to blur the edge, but the result is that most observers fail to see a well-defined surface then (Erens, Kappers, & Koenderink, 1993).

The drawing by Rimmer (1970; Figure 2, right), a fine example of an academical nude, is made up as a juxtaposition of standard stimuli, blended to a coherent undulating surface. It is an example of the technique of "ovoid drawing" (Hatton, 1904), in common use since the Renaissance. A careful study of the drawing reveals that Rimmer assumed (the drawing was most likely done without reference to a model) a light source from the top-left-front, but that he artfully changed the (local) light direction from place to place, no doubt in the interest of the overall composition, and clear delineation of muscle groups.

Such local changes in light direction are common in drawing. It is often necessary because the shading is unable to reveal "valleys" or "mountain ridges" that run along the direction of light flow (Imhof, 1965) and is one reason why realistic drawings are often preferred over photographs in anatomy or cartography. It is assumed that the human observer will tolerate such local changes without complaints.

Remarkably, relatively little is known regarding the ability of human observers to estimate the spatial distribution of light flow over articulated surfaces. Observers are certainly able to estimate overall light direction fairly precisely (Koenderink & Pont, 2003; Koenderink, Pont, van Doorn, Kappers, & Todd, 2007; Koenderink & van Doorn, 2003a, 2003b; Koenderink, van

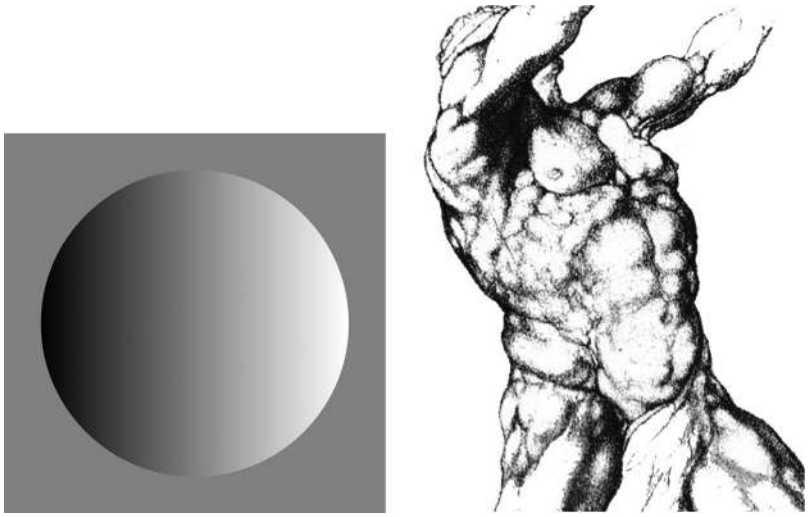

Figure 2. (Left) The "standard" stimulus configuration from psychophysics. (Right) A detail from a drawing by William Rimmer (1816–1879), a good example of "ovoid shading." Note that it is largely an arrangement of standard stimuli.

Doorn, Kappers, te Pas, & Pont, 2003; Koenderink, van Doorn, & Pont, 2004), but it remains uncertain whether observers are sensitive to spatial variations.

Human observers are also able to locate light sources in scenes even if these are occluded, that is to say, on the basis of global shading patterns (Koenderink et al., 2007). Well-known examples include Nativity scenes since the middle ages, where the Christ Child occurs as a "light source" (of divine light) at the center of the painted scene (e.g., http://collectorofechoes.files.wordpress.com/2010/02/thenativityatnight.jpg). The Utrecht School of Caravaggio followers often painted brothel scenes in which a light source (candle) was occluded by the hand or body of one of the protagonists (e.g., http://upload.wikimedia.org/wikipedia/commons/d/d6/Gerrit_van_Honthorst_-_De_koppelaarster.jpg). Since these paintings tend to be very effective, we conclude that human observers are quite likely sensitive to a divergent structure of light fields.

Various types of light fields occur frequently in the daily environment (see Figure 3). The generic possibilities are explored in some detail in Appendix A.

In this paper, we investigate the ability of observers to distinguish various types of light fields: unidirectional homogeneous fields (henceforth called "uniform"); centrally organized or "vergent" fields, where we distinguish between "convergence" and "divergence"; circularly

Figure 3. Examples of structured light fields in the daily environment. (Left) A person posed against a transilluminated curtain. This is an example of a convergent light field, the light "creeping around the sitter" as photographers say. (Bottom right) An example of a divergent light field, due to a bunch of candles. In this case, the light field fills the space; in the previous example, it flowed over the surface. (Top right) An example of an (approximately) unidirectional field. Notice that all corn grains are illuminated in roughly the same way.

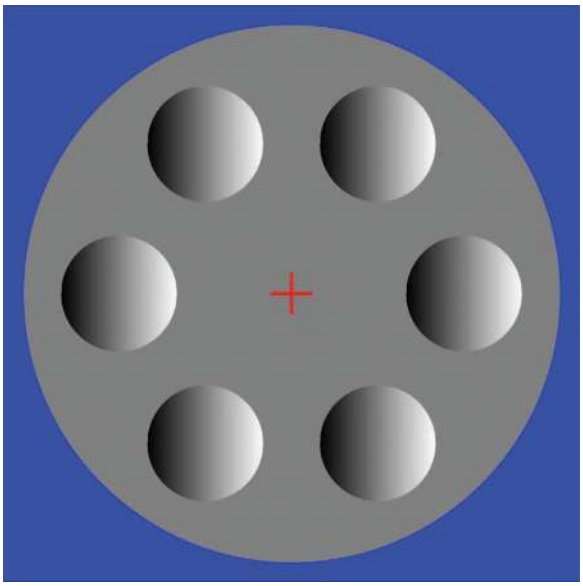

Figure 4. The stimulus pattern used in the experiments. It is composed of a circular, equispaced arrangement (at 60° intervals) of standard stimuli, placed on a circular disk of the average luminance. The disk helps combat the influence of the rectangular desktop window; we refer to it as the "pedestal." The central fixation cross was used in all cases. In this example, the light field is uniform; if the standard stimuli are assumed convex, the illumination is from the right.

organized or "cyclical" fields, where we distinguish clockwise and counterclockwise rotation; and "randomly" organized (or disorganized) fields, where the light direction varies unpredictably from place to place. In order to study these patterns, we have designed a simple, symmetrical arrangement of standard stimuli (Figure 4), emulating Rimmer's method, though without any attempt at artful "blending" (Jacobs, 1988).

The light fields used in our experiments are illustrated in Figure 5. These choices do not exhaust the possibilities (Appendix A). For instance, we decided not to include the deformation case,[2] mainly because it did not conveniently fit our experimental paradigm. The most common light fields in nature are the uniform and divergent cases, with the convergent case as a rare third (Mury, Pont, & Koenderink, 2007). Other fields (like the deformation) are extremely rare. Sunlight exemplifies the uniform case, a local source exemplifies the divergent case (this could be due to a local light spot, common on forests), and a convex object seen against a broad, diffuse source (portrait taken against translucent curtains) exemplifies the convergent case. The other cases are not ruled out by physics but require skillful laboratory setups (Mury, Pont, & Koenderink, 2009a). We included the cyclical cases as typical examples of light fields with negligible ecological importance.

Design of experimental paradigms

Stimulus and procedure

Stimulus

As one looks at the conventional stimulus for some time, one notices spontaneous reversals between convex ("cap") and concave ("cup") pictorial surfaces. As one looks at a stimulus like our array, one notices spontaneous reversals of groups (sometimes all) of the standard stimuli. Such reversals are common enough in perception, the well-known Necker cube perhaps being the most familiar (Metzger, 1975). This can be seen in Figure 4, where all standard stimuli tend to appear similar (usually caps, but

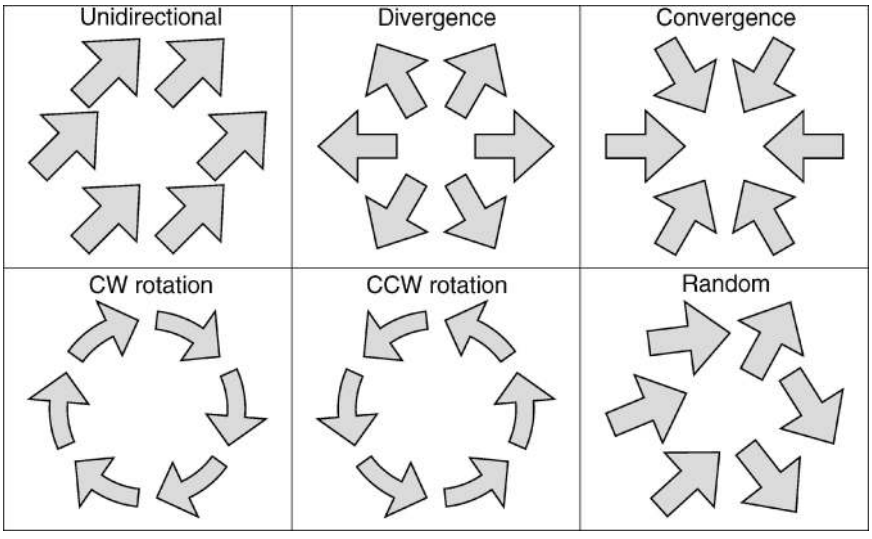

Figure 5. The light fields used in this study. Notice that the random case is only one instance out of a great many alternatives. Likewise, in the case of the unidirectional field, the direction was chosen randomly.

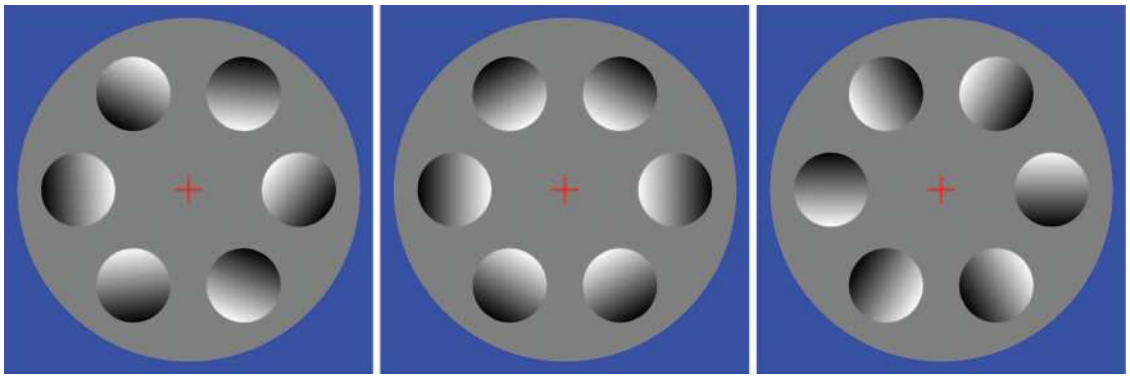

Figure 6. From left to right: random, diverging, and clockwise cyclical patterns.

sometimes cups), and in Figure 6, where this also happens for the center (diverging) case (at least for many observers). On the other hand, in the random configuration (Figure 6, left), one tends to become aware of a fluctuating pattern of cups and caps. For a cyclical pattern (Figure 6, right), one also tends to experience a mixture: a fixated disk tends to appear as a cap, whereas its antipode on the circle tends to be of the opposite type. This makes it hard to design a viable paradigm. Evidently, exposure times should be made short enough to avoid too high a frequency of spontaneous reversals. Moreover, fixation should be controlled somehow, since fixated disks tend to appear as caps.

Procedures

Eventually, we converged on two paradigms that yield different but mutually complementary observations. We indicate them as the "simultaneous" or "global" and the "successive" paradigms.

In the global paradigm, the observer is confronted with the following sequence of events:

1. In the initial condition, only the large, uniformly gray pedestal with fixation cross appears. The observer is free to choose the time to trigger the next event, after having established strict fixation of the center mark. After triggering the next event, the image does not change for a fixed (short) period (250 ms).
2. The actual configuration appears and remains on for a short period (2 s). The observer maintains fixation and notices whether either one of three mutually exclusive cases enters visual awareness: All disks appear as caps (response "convex"), all disks appear as cups (response "concave"), or a mixture of caps and cups appears (response "different").
3. After the set period of 2 s, the six disks disappear and the observer is left with the pedestal with fixation cross. In the interface, a group of three radio buttons marked "convex," "concave," and "different" appears. The observer selects the appropriate one and next hits the "done" button, which concludes the trial. The duration of this period is up to the observer.
4. The system has returned to the initial condition, and the observer may take a rest or trigger the next trial.

Thus, the observer has to attend to the pattern as a whole, which is why the paradigm is called "global."

In the successive paradigm, the fixation mark is also present. The observer is confronted with the following sequence of events:

1. In the initial condition, only the large, uniformly gray pedestal with fixation cross appears. The observer fixates the mark and triggers the next event.
2. The actual configuration appears and remains on for a short period (1 s). The observer keeps on fixating. This period is used to allow the generation of a visual awareness of the pattern.
3. The configuration stays on, but a red mark appears in one of the disks so as to mark it as the *first target*. The observer is supposed to remember the shape (cap or cup) of the marked disk. This period is only short (500 ms).
4. The configuration stays on, but the first red mark disappears and a second blue mark appears in one of the disks so as to mark it as the *second target*. The observer is supposed to remember the shape (cap or cup) of this disk. It may occasionally happen that the first and second targets are the same. This period is again a short one (500 ms).
5. After the set period, the six disks (as well as the second mark) disappear and the observer is left with the pedestal with fixation cross. In the interface, two groups of two radio buttons marked "convex" and "concave" appear. The groups are marked "first" and "second." The observer selects the appropriate button in each group and next hits the "done" button, which concludes the trial. The duration of this period is up to the observer.

6. The system has returned to the initial condition, and the observer may take a rest or trigger the next trial.

In this paradigm, the observer has to attend to two of the disks in succession, which is why we denote the paradigm "successive."

The choice of the duration of the various periods is important and was decided on the basis of pilot experiments. The observer needs some time to develop visual awareness of the pattern, yet spontaneous reversals should be minimized. We experimented with a number of obvious variations of these paradigms; the present choice seems optimal to us. It is easily possible to conceive of variations that would allow one to collect more detailed information from a trial, but we found that this renders the task an impossible one for most observers. The paradigms described here allow for hundreds of trials to be completed in a session of an hour. We did many trials in order to ensure statistically reliable results.

Experiment

Methods

Procedure

The stimuli were presented on the LCD screen of a Macintosh notebook (332-mm width screen, 1400 × 900 pixels, Macintosh LCD screen profile) in a darkened room. The observers were fully aware that they were looking at a computer screen. The user interface was presented on the same screen. Observation distance was 50 cm, and viewing was binocular. The pedestal subtended 11.4° of visual angle.

Observers were members of the different laboratories who volunteered to participate. Fourteen participated in the global paradigm, and eight participated in the successive paradigm. Two of them were authors, and the remaining observers were naive with regard to the details of the methods and the goals of the study. In addition, some others were given a few practice trials, but they were not tested formally because they appeared unable to reach satisfactory "monocular stereopsis" (i.e., three-dimensional spatial vision on the basis of monocular cues) at all.

Observers were instructed by way of a few demo trials presented by the instructor. The (simple) interface was explained to them and the instructor made sure the observers understood both the task and the interface. Trials were randomly generated and data collection continued until the observer was ready to quit and the total data volume was at least 500. For the global paradigm, the median number of trials in a session was 734 (inter-quartile range: 580–800). For the successive paradigm, the median number of trials was 558 (inter-quartile range: 502–605).

Whenever statistical tests were needed to support the claims below about differences between conditions or deviations from either no correlation (correlation 0) or perfect correlation (correlation 1), we used non-parametric bootstrapping as a technique (see Efron & Tibshirani, 1993). In short, we started from the raw responses per trial, per observer, and per condition, and resampled these 10,000 times with replacement. Using these empirically simulated distributions (rather than theoretically assumed distributions in parametric tests), we then calculated simple pairwise differences and confidence intervals. In case of multiple comparisons on the same sample, we used Bonferroni correction for standard $p < 0.05$ (unless reported otherwise).

Results

Results from the global paradigm

In the global paradigm, we collect fractions of "convex," "concave," and "different" responses for all cases. The response times were about 1 s (median value: 1112 ms, inter-quartile range: 752–1277 ms). Since the fractions add to one, the responses can be summarized conveniently in the form of pie charts. In Figure 7, we show the mean data over all observers ($N = 14$).

The results are clear cut with respect to the random and uniform categories. For the random category, almost all responses (86%) are "different," much as expected. In the case of coherent responses, we find twice as many "convex" as "concave" responses (9 and 5%, respectively). For the uniform category, only 7% of the responses are "different." There are many more "convex" than "concave" responses (68 and 25%, respectively, thus almost three quarters (73%) of all convex or concave responses are convex). One might expect a fifty–fifty distribution here since all light directions occur equally. Apparently, observers are strongly biased toward caps as opposed to cups (Langer & Bülthoff, 2001). There is quite a wide spread over the group of observers though, some being quite symmetric with respect to cap or cup responses, others reporting essentially only caps (Figure 8). No observer had a preference for reporting cups.

The responses for the cyclical categories are not significantly different from those for the random category: 88% different, 8% convex, and 4% concave for the clockwise rotation and 85% different, 10% convex, and 5% concave for the counterclockwise rotation.

The vergence categories reveal a very significant difference for the category of divergence and that of convergence. For the category of divergence, we find 41% different, 57% convex, and 2% concave responses. The category of convergence has a very different convex–concave ratio; we find 18% different, 37% convex, and 45% concave responses (all $p < 0.0001$).

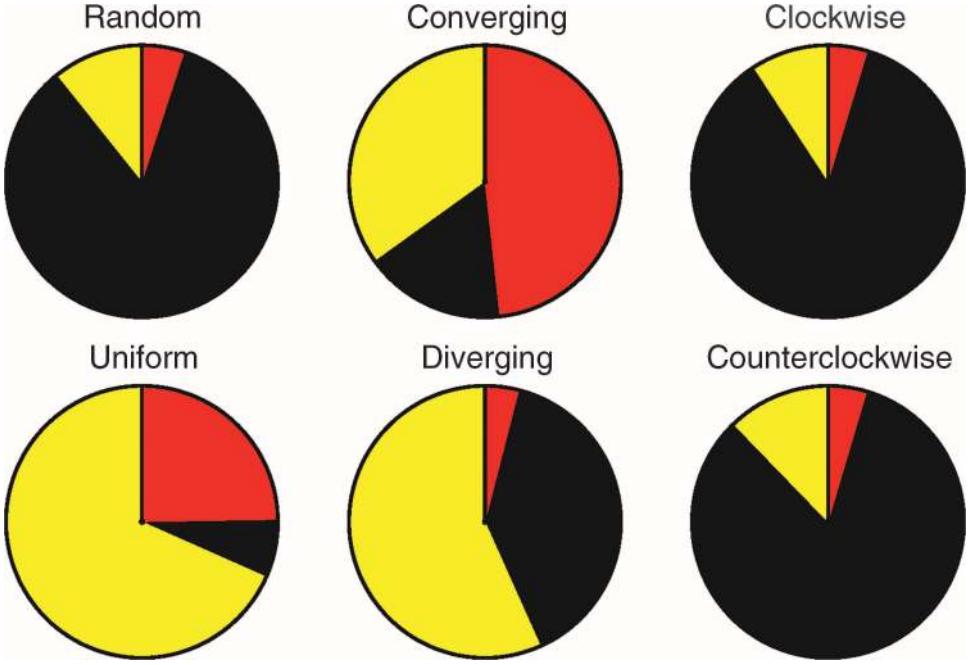

Figure 7. The mean responses over all observers. The meaning of the colors is given as follows: Black: different; yellow: convex; red: concave.

For the random, uniform, and cyclical categories, we find that observers behave similarly, but for the vergence cases, we find very significant inter-observer variations.

For the uniform category, we can check the classical preference for "light from above" (Metzger, 1975). That is to say, one expects standard stimuli that are light on top and dark at bottom to appear convex, whereas one expects the opposite orientation to look concave. This is indeed the case for almost all observers (Figures 9 and 10). We defined an index U that is zero for the case of no preference.[3] Positive values imply a preference for light from above, and negative ones imply a preference for light from below. There were only two observers in the latter category. The median index value was 1.435, a very significant preference for light from above, and the interquartile range is 0.707–2.80, with the extremes −0.771 and 3.138. Thus, there is a rather broad spectrum of preferences among our observers.

The light from above and the convex over concave preferences are distinct preferences. In this study, we essentially average out the former, but we cannot neutralize the latter preference.

The light from above preference is actually slightly shifted toward the left, as has been suggested various times in the literature (Adams, 2007, 2008; Adams, Graf, & Ernst, 2004; Mamassian & Goutcher, 2001; Sun & Perona, 1998). In order to determine this deviation, we calculated the first-order terms of the Fourier spectrum and determined the phase from that (Figure 10, right). For the observers with a U index above the median value, the median direction indicated a slight preference for directions toward the left, 16° for the convex, and 6° for the concave responses.

Results from the successive paradigm

In the case of the successive paradigm, we can find the probability of the first and second targets being seen as the same (cap–cap or cup–cup, both coded as +1) or different

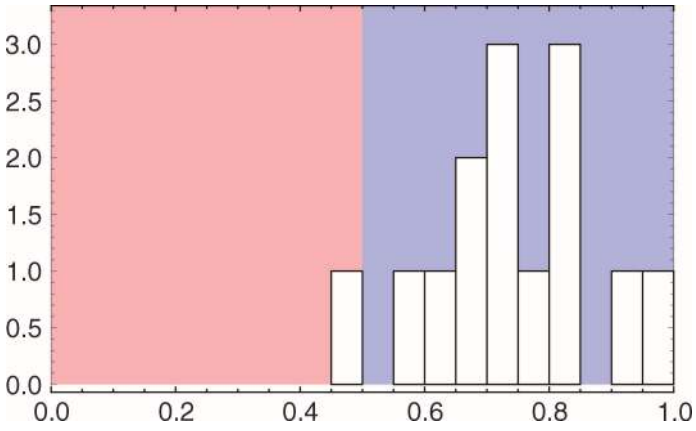

Figure 8. Histogram of the fraction of convex responses over all observers for the uniform category. For a fraction 0, observers would respond all cups; for a fraction 1, all caps; for a fraction 0.5, fifty–fifty caps–cups. The median value is 0.71, quartile range is 0.68–0.82, and extremes are 0.496 and 0.973.

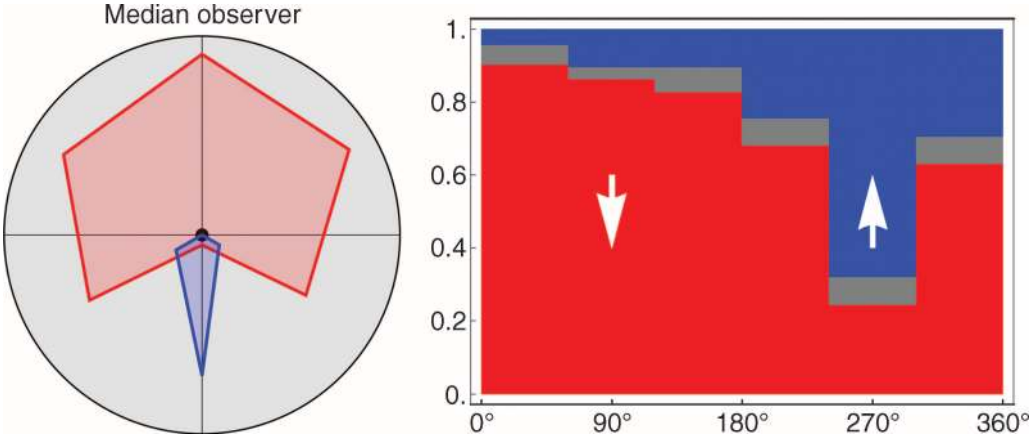

Figure 9. The fraction of "all cap" (red) and "all cup" (blue) responses as a function of the orientation of the gradient. In the left-hand plot, the top represents "light from above." This plot is for the "median observer." On the right, we show the full data, including the "different" responses, plotted in gray.

(cap–cup or cup–cap, both coded as -1) as a function of the mutual angular distance of the targets. That is to say, we can summarize the data through the autocorrelation function[4] over the circle of targets (i.e., mean values across the same or different pairs). This autocorrelation function over the pooled data is shown in Figure 11 (all observers, $N = 8$).

The self-correlations (first and second targets identical) are all very near unity (i.e., never significantly different from 1), the median value is 0.992, and the quartile range is 0.983–1. This indicates that the probability of a spontaneous reversal during a trial was very small and can be largely ignored in the analysis (but see below).

The autocorrelation functions for the various categories are qualitatively different from one another. The random and uniform categories again yield the most clear-cut results. In the case of the random category, the autocorrelations are never significantly different from zero (very small values of either sign) except for the self-correlation, which is equal to one (i.e., never differs from 1). Thus, the targets at different locations are fully independent, irrespective of their mutual distance. This is, of course, much as expected. For the uniform category, all values are very significantly positive (i.e., always significantly different from 0 at $p < 0.0001$; median value: 0.90, interquartile range: 0.86–0.94). Thus, the response to the first target has a high predictive value for the response to the second target independent of the mutual distance. This clearly reflects the informal observations.

The case of the cyclical categories is qualitatively different. Here, the correlations are generally low (except for the self-correlation) and of both signs, but—different from the case of the random category—they show a clear pattern: nearby locations are positively correlated, whereas far (roughly antipodal) locations are negatively correlated (or anti-correlated). For the counterclockwise

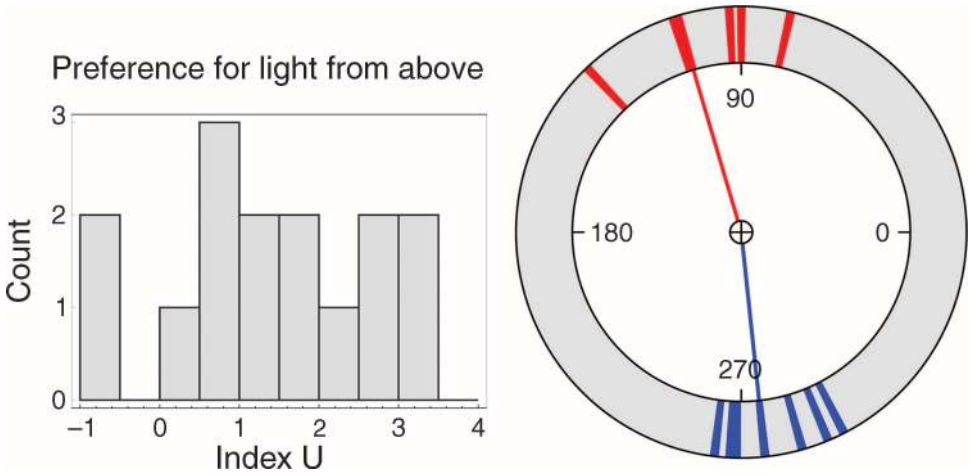

Figure 10. (Left) A histogram of the preference for light from above index *U*. (Note that negative values imply a preference for light from below; 1 count means 1 subject.) (Right) The preferred directions of the convex (red) and concave (blue) responses, as determined from the first-order Fourier components, for the observers having a *U* index above the median value.

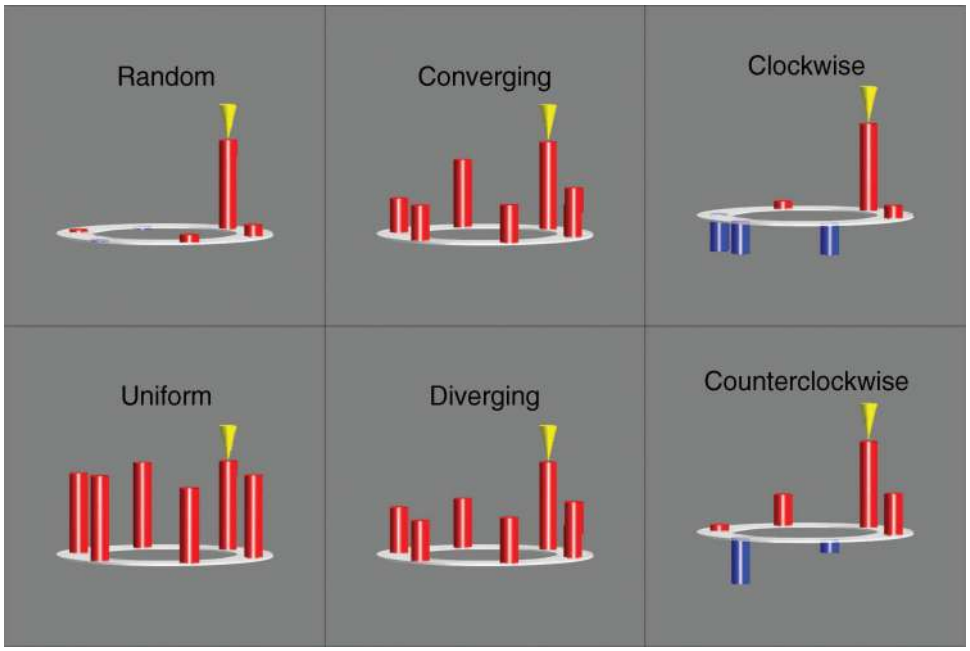

Figure 11. The autocorrelation function over the pooled data (all observers, $N = 8$), split with respect to category. The height of the bars denotes the correlation with respect to the location indicated by the yellow arrow. The tip of the yellow arrow indicates a value of 1, the red bars denote positive correlation, and the blue bars denote negative correlation.

condition, this is the case for the exact antipodal position only ($p < 0.0001$); for the clockwise condition, this is true for the antipode and its two neighbors (all $p < 0.0001$). This again reflects the informal observation (Figure 6, right) that the disk that is opposite to the fiducial one tends to be of opposite type (cap or cup as the case may be).

In the case of the vergence categories, we find very significant positive correlations, roughly uniformly distributed, only slightly dependent on mutual distance. The correlations are lower than for the uniform case though: for the divergence case, we have a median correlation of 0.53 (inter-quartile range: 0.49–0.56) and for the convergence case, a median correlation of 0.42 (inter-quartile range: 0.39–0.59). All these correlations are significantly lower than those in the uniform condition ($p < 0.005$), except at the fiducial position where none of the values differ.

The values of the correlations are shown in Figure 12. Here, the two cyclical categories have been pooled, as there is no particular reason (unlike the vergence categories) to expect them to be different. The differences between observers are strongest for the vergence categories.

Further analyses of the results

There are a number of mutually related issues to consider. We will analyze them separately in this section and only attempt an integration in our conclusions.

Rate of spontaneous reversals

The rate of spontaneous reversals is clearly very low, given that the inter-quartile range of the autocorrelation for zero distance is 0.983–1. It implies a probability of at most 0.0085 to encounter a spontaneous reversal within a 1-s interval (the total duration of the two intervals that have the markers) in case caps and cups reverse with equal probability.[5] The half-life of a certain awareness (cap or cup) then would amount to at least 82 s.[6] This seems reasonable given our observations. Informal observations appear to suggest a much shorter half-life time,

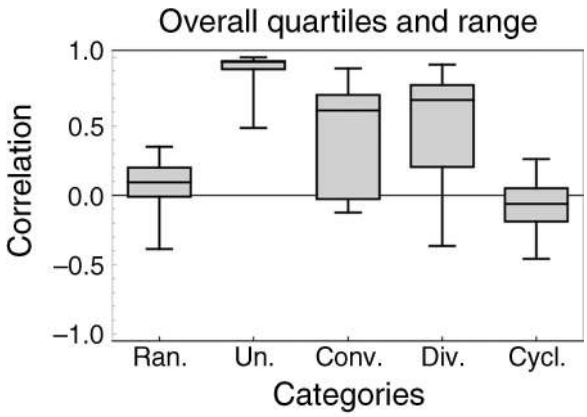

Figure 12. The values of the correlations (ignoring the self-correlation) for the various categories, with the variation between observers. The two cyclical categories have been combined.

but this is no doubt due to eye movements. With strict fixation, the spontaneous reversal rate is much slower. Note, however, that the estimate of the rate of spontaneous reversals is rather uncertain. The lowest value of the self-correlation is 0.8333, implying a rate of 0.0833 s^{-1} (a half-life time of 8.3 s), which is an order of magnitude larger than the earlier estimate of 0.0085 s^{-1}.

Light fields as entities in visual awareness

Although light fields can be estimated successfully in actual scenes (albeit with some marked exceptions), observers are typically only subsidiarily aware of them (Koenderink et al., 2007). For instance, light fields are rarely mentioned in verbal accounts of scenes, except in the context of the visual arts (photography, painting, …). One usually makes do with an overall appraisal of the "luminous atmosphere" ("sunny day," "gloomy sky," …). In psychological research, such integrative frameworks are detected as tendencies of spatially separated entities to somehow synchronize visual qualities. They often come under headings like "common fate" (Metzger, 1975) and "spatial integration" (Palmer, 1999). In our experiment, we find striking examples of such integrative frameworks that may be taken as an operational definition of "light fields."

In the case of the random configuration, the six disks are rarely seen as all cups or all caps, whereas if each disk was shown in isolation, it would more probably be seen as a cap. Let the probability of any disk in isolation to be seen as a cap be Q. Then, the probability to see all six disks as the same (either cup or cap) is $P = Q^6 + (1 - Q)^6$.[7] Empirically, the value of Q (from the mean over all observers) is approximately 3/4, implying $P = 0.1782$. Empirically, we find $P = 0.136$, which would correspond to $Q = 0.717$. Given the spread in individual observer data, this is close enough to conclude that the six disks are essentially independent in visual awareness. This is, of course, fully corroborated by the autocorrelation function, which is not significantly different from zero for finite separations.

In the case of the uniform configuration, we observe a completely different pattern. Observers tend to see all cups or all caps with high probability (0.9291). Suppose the "ideal" response would be all caps or all cups, subject to deterioration due to spontaneous reversals occurring during the trial. Here, "ideal" simply means that such a response would have "explained" the input perfectly. Assume that spontaneous reversals might start at any of the locations. Then, one would expect from the rate of spontaneous reversals to find a probability of 0.9745 if caps and cups are assumed to flip with equal probability of 0.0085 s^{-1}. The probability of 0.9291 would imply a spontaneous reversal rate of 0.0236 s^{-1}.[8] Given the uncertainty in these estimates (see above), these values appear reasonable.

The median value of the autocorrelation is 0.90, which is somewhat lower than the self-correlation that is close to 1.0. Apparently, the probability to see two different entities as equal (cap–cap or cup–cup) in the uniform configuration is about 0.95. This is much larger than an estimate on the basis of independency, which is 0.625.[9] Apparently, there is a strong tendency to "synchronize" in the case of the uniform light field.

We apparently see a tendency for the "ideal" response, perturbed through the spontaneous reversal mechanism, to which each location is subjected individually. Here, we assume the system to move closer to an ideal state soon after a reversal (which might be "all caps" instead of "all cups" and vice versa), implying that a spontaneous reversal at one location might perhaps lead to reversals at other locations. This is certainly the subjective impression.

The two vergence configurations are rather different. This asymmetry essentially rules out linear models because formally the two vergence cases differ only by a sign. Apparently, there is a preference for divergence over convergence. Notice that the convergence stimulus with caps can just as well be interpreted as a divergent one with cups. Thus, the convexity preference and the divergence preferences are in constant interaction. The asymmetry between the vergence cases can possibly be traced to the convex over concave preference.

In the case of the cyclical configurations, it seems clear from the autocorrelation function that the visual system attempts to "explain" the input by way of a uniform light field. Since there can be no global success, the configuration splits up into two opposite uniform halves. This is evident from the fact that the correlation for antipodal locations is negative. However, remarkably, this happens only in the rotation case, not in the otherwise very similar vergence cases. A linear mechanism is unable to explain this. We take it as a strong indication that the visual system uses a form of template matching, having templates for uniform and vergence configurations but not for rotations. We explore a simple example (we opted for utmost simplicity rather than data fitting in order to maximize the conceptual power of the formalism) of such a model in Appendix B. The model results indicate that such a template model is at least possible and coherent.

Differences between the categories

In Figure 13, we plot the coherency for all categories. The coherency is defined as the fraction of "all cup or all cap" responses minus the fraction of "different" responses. It varies between minus one (all different) and plus one (either all caps or all cups). Not surprisingly, the uniform configuration leads to the highest coherency, close to the theoretical maximum of +1, whereas the random configuration leads to the lowest coherency, close to the theoretical minimum of −1. These cases can be regarded as paradigmatic.

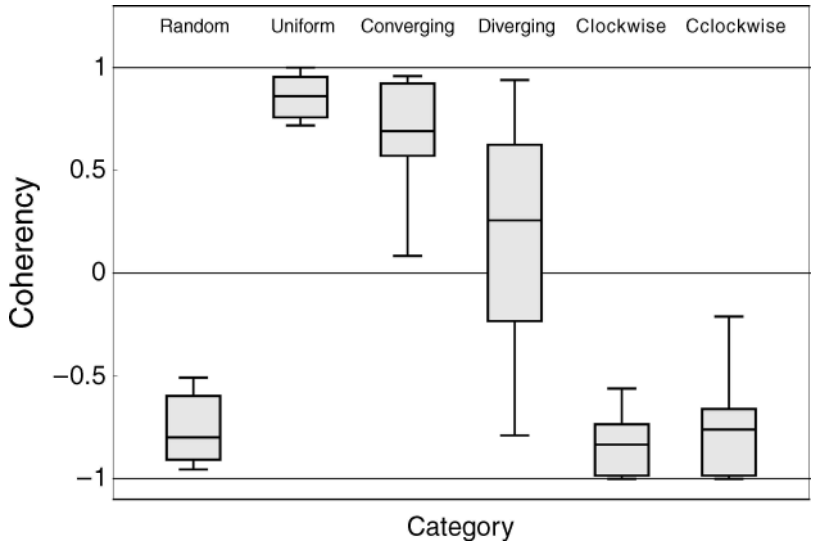

Figure 13. The coherency (the whisker–box plot shows median, quartile range, and extremes) for all categories pooled over all observers.

The cyclical configurations evidently fall in the same class as the random one; the coherency is as low as it can be. The vergence configurations are in between, though the converging configuration is actually very close to the uniform one, except for a single outlier. The diverging configuration is really "in between," its extreme values spanning much of the theoretical full range. This is the category where very significant inter-observer variations are found.

A more detailed insight is obtained by summarizing the pie charts (Figure 7) as points in a parameter space, by using the triple of fractions (always adding to the full disk) as barycentric coordinates in an equilateral triangle (Figure 14).

The vertices of the triangle are "all caps" (bottom right), "all cups" (bottom left), and all "different" (top). Since any pie chart maps on a point in the interior of the triangle, this enables one to take a global view of groups of such pie charts. For instance, one may plot the results of all observers for a certain category. The resulting point cloud yields a convenient overview of all responses for that category. We summarize the results even more conveniently by representing the point clouds with quartile and median contours. Some examples are shown in Figure 15.

In Figure 15 (left), we use this method to display the responses for the random and uniform configurations for the global paradigm and all observers. Notice that the regions are very well separated. The uniform region is located near the horizontal, lower edge, indicating very high coherency. It is located at a convexity of about 0.5, indicating a very pronounced preference of convex over concave responses. In contradistinction, the region for the random configuration is located near the top vertex, indicating very low coherency, though somewhat elongated toward the location of the region for the uniform category. This indicates that the preference of convex over concave is still in effect.

In Figure 15 (middle), we plot the region for the combined clockwise and counterclockwise cyclical configurations. The region looks much like that obtained for the random configuration, though there seems to be a slightly greater preference for convexities.

The most interesting plot is that for the vergences (Figure 15, right). These regions have very different locations, and very different shapes, in the sense that the directions of elongation are markedly different. In the divergent case, responses are predominantly convex, and

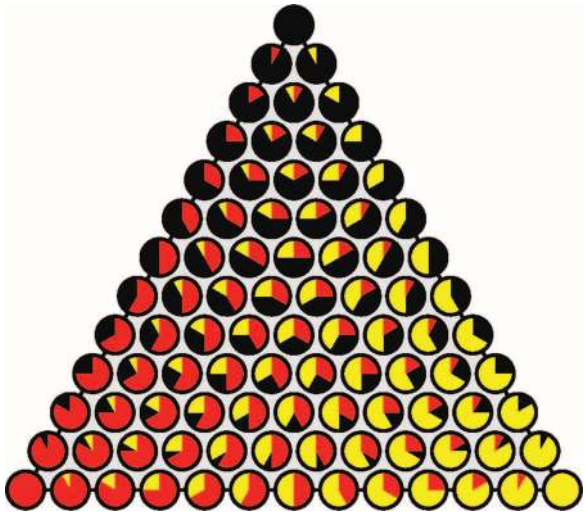

Figure 14. Barycentric coordinates allow one to represent a pie chart (as in Figure 7; same color scheme used) as a point in the interior of a triangle. This enables one to take a global view of many such pie charts simultaneously.

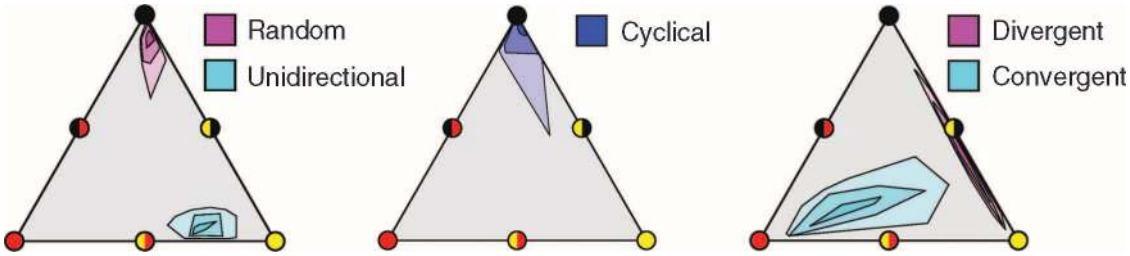

Figure 15. Inter-quartile regions and medians for the distributions of observer responses in the cases of (left) the random and uniform configurations, (middle) the cyclical configurations, and (right) the vergence configurations.

the elongation is in the direction of increased incoherency. In the convergent case, the responses are predominantly concave, and the elongation is in the direction of increased convexity. This is very interesting because it implies that in the convergence configuration the preference for convexity is relaxed and overridden by a preference for divergence. If the observer treats the disks as concave, the light comes apparently from the inside, not from the outside. One may speculate that in cases where convexities are reported, the light field is *blackness* from the inside, but this is not something we can test on the basis of the data.

Inter-observer variability

Inter-observer variability can be judged from Figures 8, 10, 12, 13, and 15, from the numbers given for the "light from above" preferences. It is evidently appreciable, though not enough, to mask a number of important trends. It is difficult to analyze these variations in detail. There is no doubt structure in them, variations being much more pronounced into some direction than into others. Here, we are handicapped by the fact that some observers seem to be unable to reach satisfactory "monocular stereopsis" at all. In this respect, our sample is biased, because we decided not to run persons in the experiment that appeared (from a cursory examination) to fail in monocular stereopsis. We (very roughly) estimate that the fraction of such cases might be as large as one in five.

Conclusions

This investigation yields a number of unexpected findings. Moreover, it corroborates some established beliefs and perhaps suggests that others might be less clear cut than usually made out.

The "light from above" preference (Metzger, 1975), mention of which never fails in even popular accounts of the shading cue (Baxandall, 2005), is very evident in the data for the uniform configuration. It is very outspoken in many observers (e.g., Figures 9 and 10), though hardly present at all in some. There evidently exists a broad spectrum of preferences.

We find that an array of conventional stimuli (Figure 2, left) tends to be perceived as all concave (or convex) if all lined up. This is already known from Ramachandran's (1988a, 1988b) demonstrations; the data presented here merely put it in a quantitative, objective format. What is new, and perhaps surprising, is that this also applies to divergent light fields and to convergent light fields but not at all to cyclical light fields. A reverse gradient in a field in the company of mutually lined up gradients tends to look opposite (e.g., a cup, as the majority is likely to look caps). However, in the context of a divergent light field, there are many opposite gradients, yet all of them look like caps. Perhaps surprisingly, this does not work out the same way for cyclical arrangements, for then the percept tends to split up, the fixated item tends to be a cap, whereas simultaneously its antipode would be a cup.

These findings are very constricting for formal accounts. We present a simple interpretation in Appendix B, based on a template model. The model, even in its—intentionally—most simple form captures the structure of the data qualitatively and even semi-quantitatively quite well. To model this behavior in a more generally appreciated manner (most probably a neural network emulating a driven Ising model; Ising, 1925) is likely to involve an overdose of ad hoc tweaks. We conclude that the data can, by and large, be understood in terms of only a few principles, namely, the template matching scheme, the convex–concave asymmetry, and the generation of spontaneous reversals.

In order to make progress beyond what we have presented here, we believe that it would be absolutely necessary to study a much larger group of observers. This will be required because of the rather large inter-observer variability. It implies a major undertaking though. Such variability is very useful because it yields a valuable handle on the structure of the underlying mechanism. One should be able to trace inter-observer variability to different

"settings" of a few parameters of a formal model. The present work has offered the first empirical and conceptual building blocks of such a more extensive endeavor.

Appendix A

Ecology of illumination patterns

"Generic" in the mathematical sense means "structurally stable" (Poston & Stewart, 1996), that is to say, small changes in conditions do not lead to qualitative changes. One expects to encounter generic phenomena as a matter of course, whereas non-generic phenomena occur once in a lifetime.

This is why it makes sense to limit the discussion to generic cases. (It is similar to the geometer's use of "in general position." Two points on a line may be assumed to be distinct, and they coincide only "once in a lifetime.")

In discussing generic illumination patterns, one needs to distinguish a few categories. Most important are:

1. the light field (Gershun, 1936/1939; Moon & Spencer, 1981), a flow in three-dimensional space;
2. the surface illuminance flow, a flow over (two-dimensional) surfaces (Pont & Koenderink, 2003, 2004).

Moreover, one needs two subcategories of the first category:

1. the light field due to a point source ("collimated beam");
2. the light field due to arbitrarily extended sources ("diffuse beam").

The first category is standard in simple computer graphics and formal discussions of "shape from shading," and the second is the typical choice of professional photographers and portrait painters.

In discussing generic illumination patterns, one also needs some basic radiometry (Boyd, 1983). The simplest setting implies Lambertian surfaces and the absence of multiple scattering. It is common in the literature on shape from shading in computer vision and experimental psychology. Corrections for more general cases are well understood (Koenderink & van Doorn, 1983, 1996, 1998a, 2003a, 2003b; Koenderink et al., 2003, 2004).

For Lambertian surfaces, "shading" can be traced to local surface irradiance. Both viewing and illumination geometries are irrelevant once the irradiance is known. This is why the "Lambertian assumption" is the obvious first approximation. For collimated beams, the irradiance is given by Lambert's cosine law; in cases where the beam is parallel (in computer graphics jargon "point source at infinity"), the irradiance is proportional to the cosine of the angle subtended by the surface normal and the direction of the beam. For general (diffuse) beams, this does not work. One defines the "light vector" such that Lambert's cosine law pertains to the light vector instead of the beam direction. The light vector describes the average transport of radiant flux. The *field lines* of the light field then replace the *rays* of the collimated beam. The difference is important. In empty space, the rays are straight lines, whereas the field lines are generally curved and may even be closed (Mury et al., 2007, 2009a).

The field lines of the light field figure prominently in the jargon used by lighting professionals. Thus, professional photographers speak of light "creeping around" an object (Figure 3 (left) is a case in point), something that appears nonsensical in a pure ray description. The light field is what counts, and only in rare cases do the field lines approximate "rays." Perhaps unfortunately, this is rarely understood in computer vision and experimental psychology. An illustrative example involves a Lambertian sphere illuminated by a luminous half-space (the overcast sky approximates this case; see Figure A1). Because the sphere itself occludes the source, the light field is due to *both* the source and the object. Both direction and magnitude of the light field vary over the surface of the sphere. All of the surface is illuminated, except for the single point where the surface does not face the source. This is in contradistinction to a sphere illuminated by a parallel, collimated beam. In that case, half of the object is in shadow (think of the moon illuminated by the sun). This is exactly the reason why most portrait photographers prefer extended sources and avoid collimated beams. Computer graphics "solves" this problem with collimated beams by adding an "ambient component." This formally works, though only for convex objects, and at the cost of a violation of physics. It is a source of frequent misunderstandings (Koenderink & van Doorn, 1996).

The light field can be almost arbitrarily complicated. Its singular points (the points where the light vector vanishes) include cases of divergence, convergence, shear, and rotation. Because the geometry is in three dimensions, various combinations occur. Such cases can indeed be constructed theoretically, and they can be found empirically (Mury et al., 2007, 2009a; Mury, Pont, & Koenderink, 2009b; see Figures A2 and A3). The structure of the light field is an important factor in human visual awareness and of much interest in practical settings like interior architecture. Human observers are sensitive to the structure of light fields in space (Koenderink et al., 2007).

Surfaces interact with the light field in two ways that deserve to be distinguished. First, surfaces interact with the light field by way of Lambert's cosine law. This involves only the component of the light vector that is normal to the surface, and it causes the irradiance, that is to say, the shading proper. Second, surfaces interact with the tangential component of the light vector. This

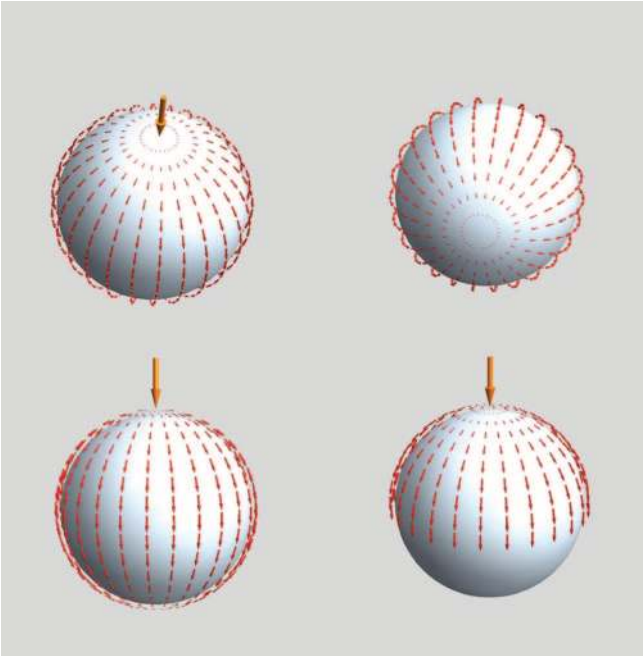

Figure A1. The light flow over the surface of a spherical object, illuminated from above (the orange arrow). For the image at bottom right, the illumination is by a collimated, parallel beam ("point source at infinity"), for all other images by an infinitely extended luminous plane ("overcast sky"). Notice that in the case of the diffuse beam all of the surface is illuminated, whereas in the case of the collimated beam, a hemisphere remains in shadow. In the former case, one has a divergent singular point (top left) and a convergent singular point (top right), whereas the typical flow pattern is uniform (bottom left). The convergent pattern is evident in the example of Figure 3 (left).

component, along the surface, causes the structure of illuminance-induced texture due to surface irregularities (Koenderink & van Doorn, 1996). Human observers are sensitive to this and perceive both the three-dimensional surface corrugations and the direction of the tangential component (Koenderink & Pont, 2003; Pont & Koenderink, 2003, 2004). We refer to the latter as the "surface irradiance flow."

The surface irradiance flow has a simple structure. An intuitive picture is the surface flow of water over a surface (Koenderink & van Doorn, 1998b). Hilltops yield diverging patterns of flow, valley bottoms yield converging patterns of flow, and passes yield shear patterns of flow. These are the only generic singular points (Guillemin & Pollack, 1974). Singular points are mutually isolated; on generic surface points, the flow is locally uniform. In radiometry, the singular points occur where the light vector hits the surface head on. On a convex object (say a sphere or an egg), one finds two singular points, one a convergence and the other a divergence (Figure A1). These are the only cases conventionally considered in the experimental psychology of shading. The convergence only occurs for diffuse sources. The standard computer graphics "point source at infinity with ambient component" approximation is insufficient to explain it, because it is inconsistent (Koenderink & van Doorn, 1996).

Thus, one has a simple overview of the generic possibilities:

1. uniform light flow is the dominating pattern, both in space and on surfaces (the corncob scene in Figure 3 (upper right) is a case in point);
2. the probability of a single object being at a space singularity of the flow is very low, but an extended configuration of objects is not unlikely to straddle such a location (the candle scene in Figure 3 (lower right) is a case in point);
3. convergence and divergence patterns in surface flows are typical (the window scene in Figure 3 (left) shows a convergent flow pattern), whereas whirl patterns are impossible.

Shears are rather less likely but possible on non-convex objects. In the case of non-convex objects that can be well approximated by convex objects ("potato-like" shapes), the shear cases will be much less dominant than the radial patterns.

In summary, one expects vision to be able to deal routinely with uniform flows, ignore rotational flows, and be able to cope with radial flow patterns and perhaps somewhat with shear patterns. Of course, none of these can be expected to occur over the full visual field but only

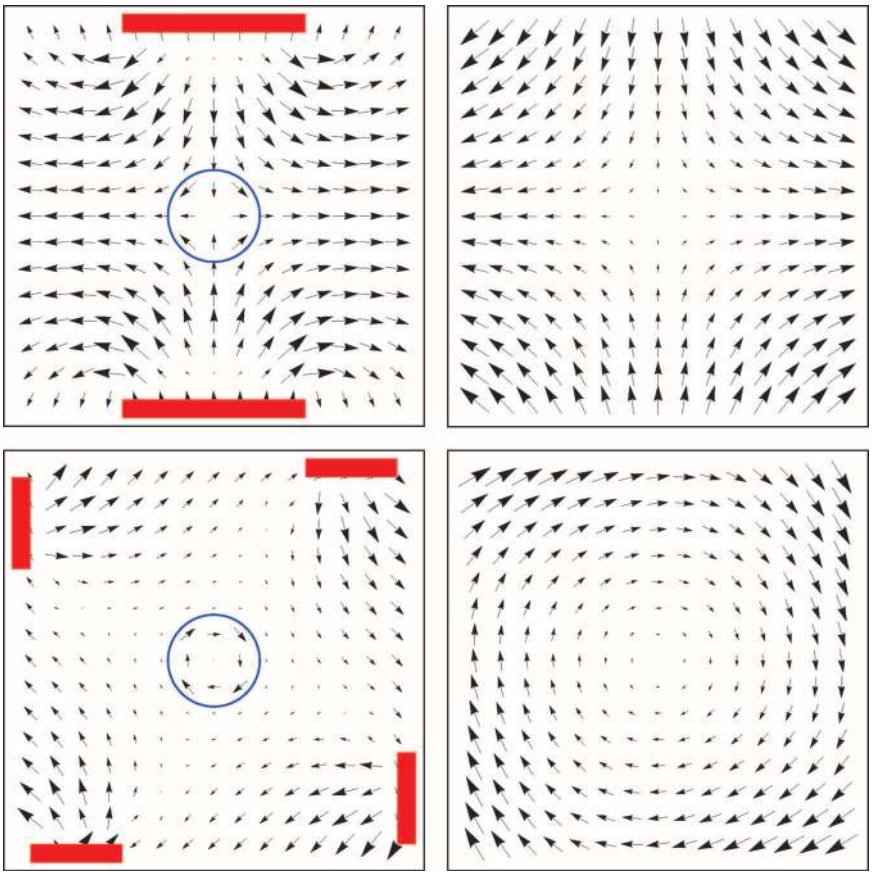

Figure A2. Two simple cases of singular points of the light field. These cases involve a cubical room, with black floor and ceiling and black vertical walls on which diffuse luminous panels are affixed, from floor to ceiling. The picture shows cross sections of the room at half-height. The luminous panels are indicated by the thick red lines. The whole room is shown in the left column, with the flow vectors in the center part magnified four times and shown separately in the right column. The top row illustrates a shear, and the bottom row illustrates a whirl pattern. Since radiometry is scale-invariant, the size of the room is not indicated. Such cases have been measured in actual situations (we used the light laboratory of Philips Company; Mury et al., 2009a, 2009b).

in specific partitions of it. For instance, the illuminated and shadow sides of a face are necessarily illuminated by different sources.

Appendix B

A template model

In this appendix, we consider a simple template-based model. The model is only designed as a conceptual aid, and we do not attempt any data fitting. We are convinced that many, conceptually diverse models might serve to explain the data in a coarse approximation. The present model is perhaps of some interest because of its extreme simplicity.

In this model, the agent essentially generates "hallucinations" and probes the front end to see whether these fit the input. Such hallucinations are by their very nature *meaningful,* the meaning being imposed by the agent. The "hallucinations" have the same format as the data in the buffer, thus the probing is a simple matter of comparison. In the model, the format is simply an ordered list of six directions, and the comparison finds the mean of the matches, a match being defined as the cosine of the relative directions, a number between +1 (same directions) and −1 (opposite directions). The agent comes up with many hallucinations and, at any moment, favors the one that fits best (Selfridge, 1959). This competition among generations of hallucinations is akin to biological "survival of the fittest." The survivors make up "visual awareness" (Hoffman, 2009). From an algorithmic point of view, the model is similar to the "harmony search" algorithm of "soft computing" (Geem, Kim, & Loganathan, 2001).

The hallucinations can be based on what is in the data buffer or can be generated in a fully autogenous manner. In the first case, the agent may pick an item from the data buffer (say the observer fixates an item) and use a vector of six identical copies as the hallucination. When

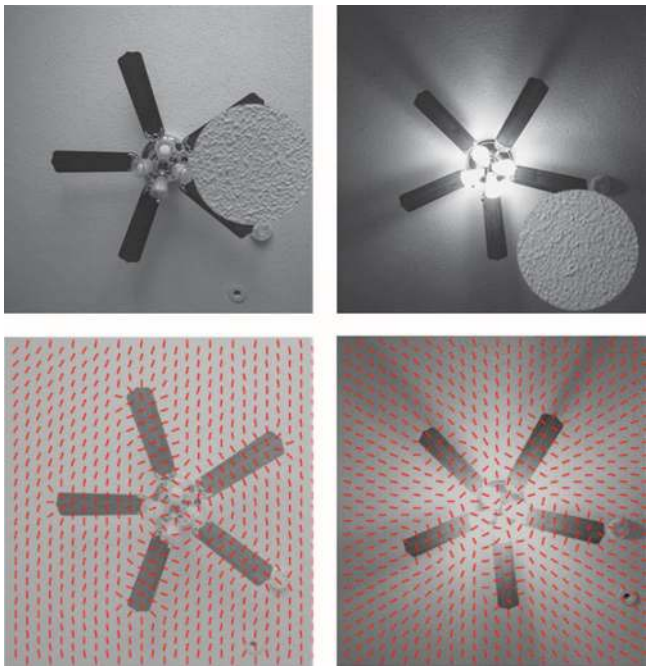

Figure A3. A ceiling ventilator lamp during (left column) day and (right column) night. (Top row) Straight photographs with a circular inset showing a four times enlargement of the texture of the plastered ceiling. (Bottom row) The raw output of a local illumination flow estimate based on second-order edge detector statistics. This is the algorithm that accounts well for human detection of illuminance flow (Koenderink et al., 2004). Ignoring the perturbations by the strong edges due to the fan blades, the daylight flow field is uniform (vertical in the image), illumination being due to daylight streaming in from a window on one side. At night, the surface illuminance flow is due to the four light bulbs and the flow is radial. Such phenomena are easy to see once one starts noticing.

generating autogenously, the agent may—in cases of minimal situational awareness—generate sextuples at random or select one from a database of templates. The selection criterion will depend on situational awareness. Such templates are akin to "Gestalts" (Lorenz, 1973; Metzger, 1975; Tinbergen, 1951; for an extensive discussion of this point, see Riedl, 1984). The database might be innate, or be the result of experience, hallucinations that frequently led to efficacious actions.

Notice that this is not a dynamical system, there being no actual coupling between the front end and the agent. The front end acts like a volatile memory, and it is continuously overwritten by the world and intermittently addressed by the agent's probing. The function of the front end is analogous to the "function" of the beach as a memory for footprints. The topography of depressions in the sand is meaningless, except to the beachcomber momentarily interested in human activity. There is no data transfer from world to awareness, meaning ("data") being internally generated. This cannot be otherwise because meaning cannot be computed from structure. It is only the constant checking against the data buffer that renders this system a viable "optical user interface." This model is biologically inspired and fits seamlessly in an evolutionary framework.

In the model, we use two types of template, one for diverging patterns and a continuous set tuned to uniform patterns of any direction. The agent tries them all and retains the one that fits best. Given a best fit, the agent constructs the response (visual awareness) from the local fits (six numbers between −1 and +1). Since fits near 0 lead to ambiguity, the agent uses a threshold (cos(45°) in the model) and interprets fits above the threshold as "convex" and those below minus the threshold as "concave." In the ambiguous cases, the observer first checks the nearest neighbors to see whether these agree (cap–cap or cup–cup), and if so conforms to these neighbors. In case of remaining ambiguity, the agent finally flips a coin. Thus, any stimulus as seen from the perspective of any hallucination yields a response that is a vector of six "cap" or "cup" decisions (no ambiguity in the response). Notice that there is some non-trivial dynamics in the response forming act, the tendency to conform to the neighbors, reminiscent of the Ising model. There are two sources of indeterminacy in the model, one in the mapping of the stimulus on the front end (an additive noise source) and one in the response forming act where the agent occasionally needs to flip a coin.

Thus far in the discussion, we have not incorporated a mechanism for spontaneous flips nor for the empirically found convexity above concavity preference. We have implemented both at the same time by adding a final stage to the response forming mechanism. We simply flip cups

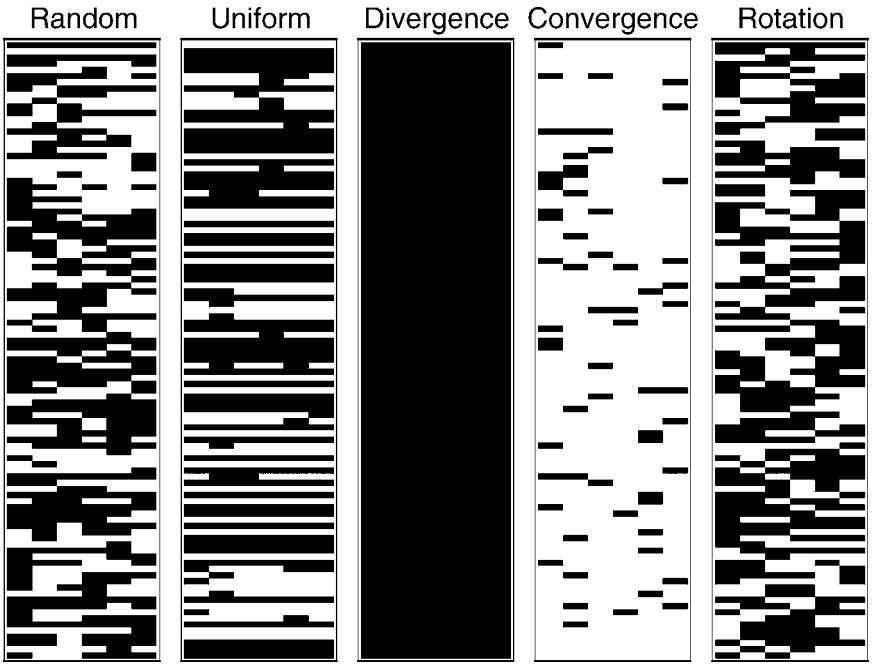

Figure B1. Examples of model responses for trials with the various configurations. Each column contains a hundred trials; each trial is plotted as a row of six abutting white or black rectangles. White indicates "cup," and black indicates "cap."

to caps with a certain probability (0.1 in the model examples). This is a third source of indeterminacy.

This simple model contains only a few parameters, the only ones being important being the level of additive noise at the input stage (we perturbed directions with a normal distribution with a standard deviation of 20°) and the spontaneous flip rate.

Obviously, the model is far too simple so as to catch all structures present in the empirical data. For instance, there is no such thing as the light from above preference in the model, since the model is fully isotropic. "Above" is simply meaningless as applied to the model.

Figure B1 shows model sessions of a hundred trials each run with the parameter values mentioned above. Notice how different the responses are for the various configurations. The responses for the random configuration are indeed random. Although the responses for the rotation appear random too, there is actually a structure in them that appears in the autocorrelation function (see below). The uniform configuration gives rise to mainly coherent responses, all cups or all caps. Even the mixed responses are not really random, there being typically only a single outlier. The divergence configuration leads to mere "all cap" responses in this session; in longer sessions, one sees occasional (single) cup outliers. The convergence configuration gives rise to essentially "all cup" responses, though with numerous "cap" outliers. The asymmetry between the two vergence cases is due to the spontaneous transitions, which are all "cup-to-cap" flips. By introducing distinct probabilities for "cup-to-cap" and "cap-to-cup" flips, one could make the model more "realistic."

The autocorrelation functions for these sessions are shown in Figure B2. They are much like the empirical results shown in Figure 11. For the random configuration, the correlations are low. They are due to the neighborhood interaction. Adding weights (or even a neighborhood weighting function) to the model would allow one to "tune" to resemble the empirical results even more. For the uniform configuration, the correlations are high. The value below unity is due to the input perturbation and to the spontaneous flip mechanism. The latter is the more important cause, as can be seen from the vergence configurations. The divergence configuration leads to *perfect* correlations because there are no spontaneous cap-to-cup flips; on the other hand, the correlations for the convergence are lower than those for the uniform case because here the cup-to-cap flips have a huge influence, most responses being of the "cup" type. The autocorrelation for the cyclical configurations are especially interesting since we see an anti-correlation peaked at the antipodal of the fiducial position. This is due to the fact that the best the model can do is apply a uniform template, with the result that the configuration perceptually splits into two parts.

There are a number of simple ways to amend the model so as to mimic the data in more detail. However, this moves one into the domain of "fitting," whereas the model is meant to be conceptual. Even with the model in its present, simplest form, we already capture the structure of

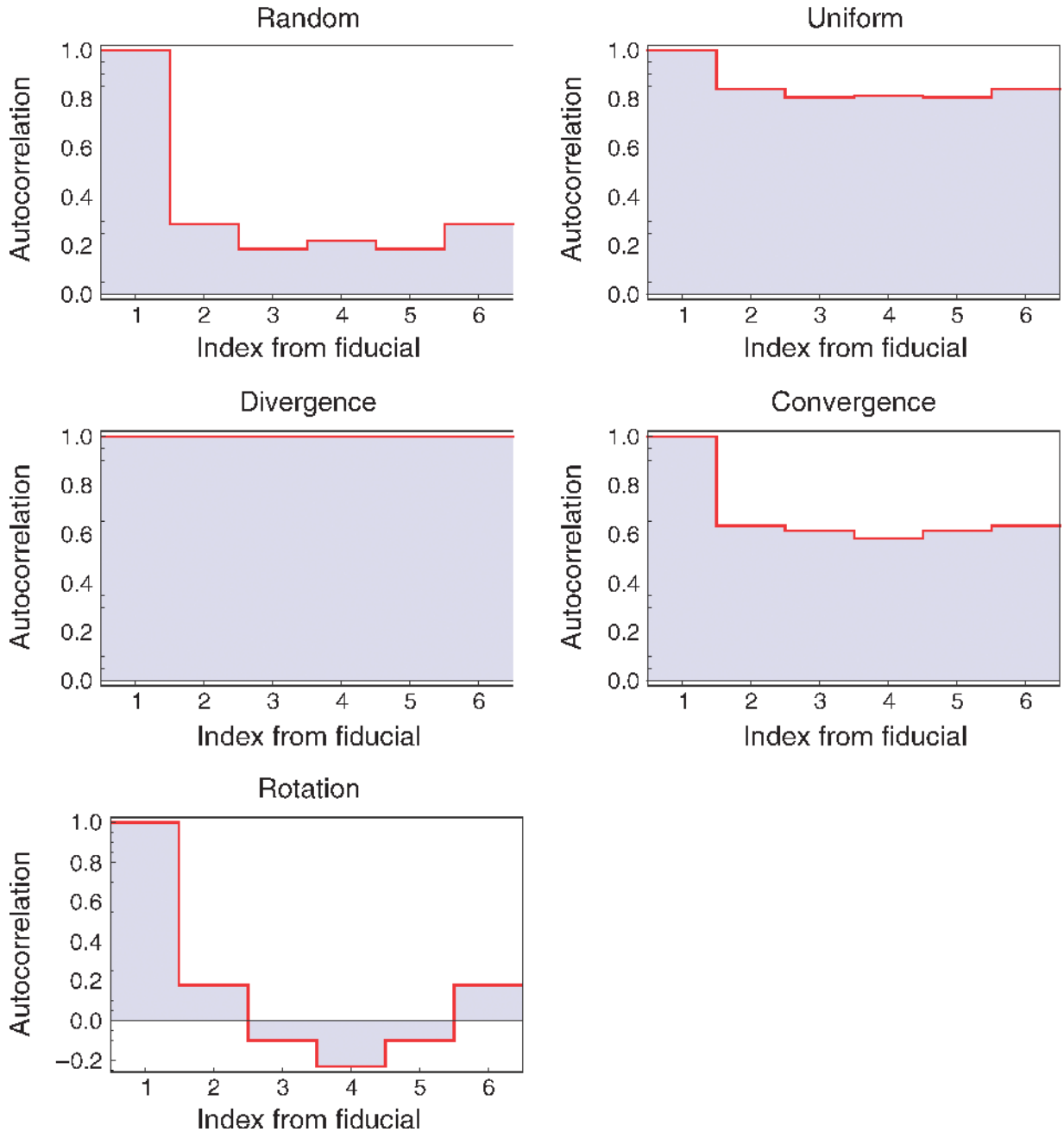

Figure B2. Autocorrelation functions obtained from model sessions of a hundred trials. The self-correlation peak is at position 1, and the horizontal axis is periodic, thus position 6 is the left neighbor of position 1.

the data quite satisfactorily. The value of the model's simplicity is then that it allows one to reason about the various trends on the basis of only a few general mechanisms.

Acknowledgments

This work was supported by the Methusalem Program of the Flemish Government (METH/08/02), awarded to JW. JK was a Visiting Fellow at the Flemish Academic Centre for Science and the Arts (VLAC) at the time of the research. We would like to acknowledge the technical support by Frank Amand, the administrative support by Stephanie Poot, the assistance with the bootstrap statistics by Tom Putzeys, and the useful advise by the reviewers.

Commercial relationships: none.
Corresponding author: Johan Wagemans.
Email: johan.wagemans@psy.kuleuven.be.

Address: Laboratory of Experimental Psychology, University of Leuven (K.U. Leuven), Tiensestraat 102, bus 3711, BE-3000 Leuven, Belgium.

Footnotes

[1]Formally, "local" structure refers to the infinitesimal domain. In applications, any structure that is described by entities defined at a point is "local". A key example is the *gradient*. Thus, "local shading" is a *point property* even though the image intensity spatially varies. It is formally similar to the notion of "velocity," which is a measure of temporal change at a single moment. In contradistinction, "multi-local properties" depend on the structures simultaneously encountered at mutually distinct points. Mathematical analysis in the local and multi-local cases has to be categorically different. Similarly, in neurophysiology, "local" may be taken to refer to the single receptive field case, whereas "multi-local" would refer to receptive field assemblies. Thus, "multi-local" presupposes local sign, whereas "local" does not. These distinctions are crucial in any formal account of spatiotemporal phenomena.

[2]Near the origin, a vector field in the plane can be developed to first order as a unidirectional, uniform field with a superimposed linear perturbation. The perturbation can be split into symmetric and anti-symmetric parts. The anti-symmetric part describes rotations. The symmetric part can again be split into an isotropic part that describes vergences and an anisotropic part that describes a pure deformation or shear. Whereas rotations and vergences are isotropic, the deformation has an orientation, the axis of elongation. Although this exhausts the generic structures, it fails to address the issue of ecological validity (see Appendix A).

[3]The data are at 60° intervals, such that 90° indicates "light from above," that is to say the top of the standard stimulus is light and its bottom is dark. Let $R^{+}_{60^\circ}$ denote the number of "convex" responses at 60°, $R^{-}_{30^\circ}$ the number of concave responses at 30° and so forth. Then, we define the "preference for light from above index U" as $U = {}^{2}\log\{[(R^{+}_{30^\circ} + 2R^{+}_{90^\circ +} + R^{+}_{150^\circ}) + (R^{-}_{210^\circ} + 2R^{-}_{270^\circ +} + R^{-}_{330^\circ})]/[(R^{+}_{210^\circ} + 2R^{+}_{270^\circ +} + R^{+}_{330^\circ}) + (R^{-}_{30^\circ} + 2R^{-}_{90^\circ +} + R^{-}_{150^\circ})]\}$. This index is essentially a ratio of "votes" for the respective categories and is symmetric in "above" and "below" preferences. A value of +1 implies twice as many votes for light from above as for light from below.

[4]The autocorrelation function $R(\varphi)$ is defined as the expectation of $F(\theta + \varphi)F(\theta)$ over $0 \leq \theta < 360°$, where $F(\theta) = +1$ for convex responses and $F(\theta) = -1$ for concave responses. Of course, the parameter is periodic, thus $\theta + \varphi$ is reckoned modulo 360°. We reckon $-180° \leq \varphi < +180°$; in the experiment, the angle φ takes only the discrete values of −180°, −120°, −60°, 0°, +60°, and +120° (where +180° repeats −180°).

[5]Suppose the *ab initio* probability to see a cap is Q, and that of seeing a cup is $1 - Q$. Empirically, $Q \approx 3/4$. Let the probability that a cap will spontaneously flip to a cup within the exposure period of 1 s be A; likewise, let the probability that a cup will spontaneously flip into a cap be A. The probability of reversal A is evidently very small; reversals are rarely noticed under conditions of strict fixation. The self-correlation R is the mean of four possible events, namely cap–cap, cup–cup, cap–cup, and cup–cap. The former two events lead to a value of +1, and the latter two events lead to a value of −1. The probabilities of the four events are $Q(1 - A)$, $(1 - Q)(1 - A)$, QA, and $(1 - Q)A$, respectively, thus one has $R = Q(1 - A) + (1 - Q)(1 - A) - QA - (1 - Q)A = 1 - 2A$. Empirically, $R = 0.983 \ldots 1$ (inter-quartile range), leading to $A < 0.0085\ \mathrm{s}^{-1}$.

[6]Assuming exponential decay $\exp(-t/T) \approx 1 - t/T$, we have $t/T \approx A$, with $t = 1$ s, thus $T = 1/A$. The half-life time is $\log 2 \approx 0.693$ times as large. One finds a half-life time of at least 82 s in case spontaneous reversal occurs symmetrically (see Footnote 5).

[7]We have six locations. Let the *ab initio* probability to see a cap be Q and that to see a cup be $(1 - Q)$. The probability to see six caps is Q^6, and that to see six cups is $(1 - Q)^6$. Thus, the probability to see six equal entities is $P = Q^6 + (1 - Q)^6$, a symmetrical function of Q, being $2^{-5} \approx 0.03125$ at $Q = 0.5$ and 1 at $Q = 0$ or 1.

[8]We have a duration of $t = 1/2$ s. If the probability of a spontaneous reversal in 1 s is A, the probability of any reversal in a period t is $6tA$. If one initially sees six equal entities, the probability to see them for the full interval is $(1 - 6tA) = (1 - 3A)$.

[9]Consider two different locations and let the probability to see two cups or two caps be S. The correlation is expected to be $R = S - (1 - S) = 2S - 1$. Empirically, $R = 0.9$, implying $S = 0.95$. Let the probability to see a single item as cap be Q, and assume the two locations are independent. Then, the probability of cap–cap is Q^2, and that of cup–cup *is* $(1 - Q)^2$; thus, the probability to see the two as equals is $P = Q^2 + (1 - Q)^2$. Empirically, $Q \approx 3/4$, implying $P = 0.625$. This is much smaller than the value of S (which is 0.95, see above).

References

Adams, W. J. (2007). A common light-prior for visual search, shape, and reflectance judgements. *Journal of Vision, 7*(11):11, 1–7, http://www.journalofvision.org/content/7/11/11, doi:10.1167/7.11.11. [PubMed] [Article]

Adams, W. J. (2008). Frames of reference for the light-from-above prior in visual search and shape judgements. *Cognition, 107,* 137–150.

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience, 7,* 1057–1058.

Baxandall, M. (2005). *Shadows and enlightenment* (2nd ed.). London: Yale University Press.

Belhumeur, P. N., Kriegman, D. J., & Yuille, A. L. (1999). The bas-relief ambiguity. *International Journal of Computer Vision, 35,* 33–44.

Boyd, R. W. (1983). *Radiometry and the detection of optical radiation*. New York: Wiley.

Brewster, D. (1832). *Letters on natural magic*. London: John Murray.

Cate, A. D., & Behrmann, M. (2010). Perceiving parts and shapes from concave surfaces. *Attention, Perception, & Psychophysics, 72,* 153–167.

Dana, K. J., van Ginneken, B., Nayar, S. K., & Koenderink, J. J. (1999). Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics, 18,* 1–34.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC Press.

Erens, R. G. F., Kappers, A. M. L., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics, 54,* 145–156.

Forsyth, D. A., & Ponce, J. (2002). *Computer vision: A modern approach*. Upper Saddle River, NJ: Prentice-Hall.

Forsyth, D. A., & Zisserman, A. (1991). Reflections on shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13,* 671–679.

Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: Harmony search. *Simulation, 76,* 60–68.

Gershun, A. (1939). The light field (P. Moon & G. Timoshenko, Trans.). *Journal of Mathematics and Physics, 18,* 51–151. (Original work published in 1936)

Guillemin, V., & Pollack, A. (1974). *Differential topology*. Englewood Cliffs, NJ: Prentice Hall.

Hatton, R. G. (1904). *Figure drawing*. London: Chapman & Hall.

Hayakawa, H., Nishida, S., Wada, Y., & Kawato, M. (1994). A computational model for shape estimation by integration of shading and edge information. *Neural Networks, 7,* 1193–1209.

Hill, H., & Bruce, V. (1994). A comparison between the hollow-face and "hollow-potato" illusions. *Perception, 23,* 1335–1337.

Hoffman, D. (2009). The interface theory of perception: Natural selection drives true perception to swift extinction. In S. Dickinson, M. Tarr, A. Leonardis, & B. Schiele (Eds.), *Object categorization: Computer and human vision perspectives* (pp. 148–165). Cambridge, UK: Cambridge University Press.

Horn, B. K. P., & Brooks, M. J. (Eds.). (1989). *Shape from shading*. Cambridge, MA: MIT Press.

Humphrey, G. K., Symons, L. A., Herbert, A. M., & Goodale, M. A. (1996). A neurological dissociation between shape from shading and shape from edges. *Behavioral Brain Research, 76,* 117–125.

Imhof, E. (1965). *Kartographische Geländedarstellung*. Berlin, Germany: Walter de Gruyter.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik A: Hadrons and Nuclei, 31,* 253–258.

Jacobs, T. S. (1988). *Light for the artist*. New York: Watson-Guptill.

Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics, 52,* 18–36.

Koenderink, J. J., & Pont, S. C. (2003). Irradiation direction from texture. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 20,* 1875–1882.

Koenderink, J. J., Pont, S. C., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2007). The visual light field. *Perception, 36,* 1595–1610.

Koenderink, J. J., & van Doorn, A. J. (1983). Geometrical modes as a general method to treat diffuse interreflections in radiometry. *Journal of the Optical Society of America, 73,* 843–850.

Koenderink, J. J., & van Doorn, A. J. (1996). Illuminance texture due to surface mesostructure. *Journal of the Optical Society of America A, 13,* 452–463.

Koenderink, J. J., & van Doorn, A. J. (1998a). Phenomenological description of bidirectional surface reflection. *Journal of the Optical Society of America A, 15,* 2903–2912.

Koenderink, J. J., & van Doorn, A. J. (1998b). The structure of relief. In P. W. Hawkes (Ed.), *Advances in imaging and electron physics* (vol. 103, pp. 65–150). New York, NY: Academic Press.

Koenderink, J. J., & van Doorn, A. J. (2003a). Local structure of Gaussian texture. *IEEE Transactions on Information and Systems, 86,* 1165–1171.

Koenderink, J. J., & van Doorn, A. J. (2003b). Shape and shading. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1090–1105). Cambridge, MA: MIT Press.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., te Pas, S. F., & Pont, S. C. (2003). Illumination

direction from texture shading. *Journal of the Optical Society of America A, 20,* 987–995.

Koenderink, J. J., van Doorn, A. J., & Pont, S. C. (2004). Light direction from shad(ow)ed random Gaussian surfaces. *Perception, 33,* 1405–1420.

Lambert, J. H. (1760). *Photometria, sive, De mensura et gradibus luminis, colorum et umbrae*. Augsburg, Germany: V. E. Klett.

Langer, M. S., & Bülthoff, H. H. (2001). A prior for global convexity in local shape-from-shading. *Perception, 30,* 403–410.

Lorenz, K. (1973). *Die Rückseite des Spiegels*. München, Germany: Piper Verlag.

Luckiesh, M. (1916). *Light and shade and their applications*. New York: Van Nostrand.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition, 81,* B1–B9.

Metzger, W. (1975). *Gesetze des Sehens*. Frankfurt a. M., Germany: Verlag Waldemar Kramer.

Moon, P., & Spencer, D. E. (1981). *The photic field*. Cambridge, MA: MIT Press.

Mury, A. A., Pont, S. C., & Koenderink, J. J. (2007). Spatial properties of light fields in natural scenes. In S. N. Spencer (Ed.), *Proceedings of the 4th Applied Perception in Graphics and Visualization (APGV 2007), ACM SIGGRAPH* (p. 140). New York, NY: ACM.

Mury, A. A., Pont, S. C., & Koenderink, J. J. (2009a). Representing the light field in finite three-dimensional spaces from sparse discrete samples. *Applied Optics, 48,* 450–457.

Mury, A. A., Pont, S. C., & Koenderink, J. J. (2009b). Structure of light fields in natural scenes. *Applied Optics, 48,* 5386–5395.

Norman, J. F., & Raines, S. R. (2002). The perception and discrimination of local 3-D surface structure from deforming and disparate boundary contours. *Perception & Psychophysics, 64,* 1145–1159.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: Bradford Books/MIT Press.

Pont, S. C., & Koenderink, J. J. (2003). Illuminance flow. In N. Petkov & M. A. Wetsenberg (Eds.), *Computer analysis of images and patterns* (LNCS 2756, pp. 90–97). Berlin/Heidelberg, Germany: Springer-Verlag.

Pont, S. C., & Koenderink, J. J. (2004). Surface illuminance flow. In Y. Aloimonos & G. Taubin (Eds.), *Proceedings of the Second International Symposium on 3D Data Processing Visualization and Transmission* (pp. 2–9). Los Alamitos, CA: IEEE Computer Society.

Poston, T., & Stewart, I. (1996). *Catastrophe theory and its applications*. New York, NY: Dover Publications.

Ramachandran, V. S. (1988a). Perceiving shape from shading. *Scientific American, 258,* 76–83.

Ramachandran, V. S. (1988b). Perception of shape from shading. *Nature, 331,* 163–166.

Riedl, R. (1984). *Biology of knowledge: The evolutionary basis of reason*. Chichester, UK: Wiley.

Rimmer, W. (1970). *Art anatomy*. New York: Dover Publications.

Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society, 2,* 37–42.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In D. V. Blake & A. M. Uttley (Eds.), *Proceedings of the Symposium on Mechanisation of Thought Processes* (pp. 511–529). London, UK: Her Majesty's Stationery Office.

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience, 1,* 183–184.

Tinbergen, N. (1951). *The study of instinct*. London: Oxford Clarendon Press.

Turhan, M. (1935). Über räumliche Wirkungen von Helligkeitsgefällen. *Psychologische Forschung, 21,* 1–49.

van Diggelen, J. (1959). Photometric properties of lunar crater floors. *Recherches Astronomiques de l'Observatoire d'Utrecht, 14,* 1–114.

Zhang, R., Tsai, P. S., Creyer, J. E., & Shah, M. (1999). Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21,* 690–706.