Annual Review of Vision Science

Geometry of Pictorial Relief

Jan J. Koenderink,[1,2,3] Andrea J. van Doorn,[2,3]
and Johan Wagemans[1]

[1]Laboratory of Experimental Psychology, University of Leuven (KU Leuven), 3000 Leuven, Belgium; email: KoenderinkJan@gmail.com

[2]Department of Psychology, Giessen University, 35394 Giessen, Germany

[3]Department of Experimental Psychology, Utrecht University, 3584 CS Utrecht,
The Netherlands

Annu. Rev. Vis. Sci. 2018. 4:451–74

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-091517-034250

Keywords

visual field, pictorial space, depth, relief

Abstract

Pictorial relief is a quality of visual awareness that happens when one looks into (as opposed to at) a picture. It has no physical counterpart of a geometrical nature. It takes account of cues, mentally identified in the tonal gradients of the physical picture—pigments distributed over a planar substrate. Among generally recognized qualities of relief are color, pattern, texture, shape, and depth. This review focuses on geometrical properties, the spatial variation of depth. To be aware of an extended quality like relief implies a "depth" dimension, a nonphysical spatial entity that may smoothly vary in a surface-like manner. The conceptual understanding is in terms of formal geometry. The review centers on pertinent facts and formal models. The facts are necessarily so-called brute facts (i.e., they cannot be explained scientifically). This review is a foray into the speculative and experimental phenomenology of the visual field.

1. INTRODUCTION

Pictorial relief is a term that describes the experience of a smoothly curved surface when looking "into" a picture, where the picture itself is a flat physical object (Koenderink et al. 2011). In this short review, we restrict the discussion mainly to relief (a surface), and not space (a volume), in static, monocularly viewed pictures (no motion parallax, no disparity) and mainly in photographs or artistic renderings of natural scenes. This implies that the larger part of the literature on depth perception in its manifold guises is intentionally ignored. Even with these restrictions, the review necessarily presents a limited scope.

Pictorial relief (Koenderink et al. 2006) is evidently unlike the perception of some surface in physical space. Maurice Denis (1890), in his *Manifesto of Symbolism*, famously declared that whatever else a picture is, it is "essentially a flat surface covered with colors assembled in a certain order" (p. 540). This insight had considerable impact on the visual arts (think of the emergence of abstract art) but largely missed an impact on vision science, which—on the whole—has yet to absorb its significance.

Denis's insight implies that the notion of veridicality fails to apply and that pictorial relief is more appropriately defined as (controlled) imagery than as perception proper. Pictorial relief exists only in visual awareness (Koenderink 2015a); thus the notion of its geometrical structure stands in need of some operational definition (Bridgman 1927) that has to be quite unlike that which may apply to surfaces in the physical environment. Various operationalizations have been implemented in vision science, and a few are discussed below. Different methods must be expected a priori to yield different results; such differences might be expected to yield additional insight.

Pictorial relief can be studied (*a*) in its intrinsic structure, (*b*) in relation to gradients of pigmentation of the picture surface, (*c*) with respect to the operationalization, and (*d*) with respect to semantic pictorial content. With respect to *d*, notice that pictorial content is essentially imagery—hence the risk of vicious circularity.

Photography does not escape Denis's dictum. Naturalistic art can yield pictures that are not spontaneously distinguished from photography by generic observers. A piece of paper found on a dung hill and overgrown with fungus might pass for a photograph; after all, both are exhaustively described as pigment distributions over flat surfaces.

Photographs sometimes come with a history—namely, the physical scene in front of the camera and the optical properties of the lens. Visual relief may be related to that (see Vishwanath et al. 2005). For most photographs seen in daily life, such a history is not available. In this (which is the most frequent) case, there can be no notion of veridicality.

1.1. Formal Structure of Relief

A formal description of relief (Koenderink & van Doorn 2012) distinguishes a substrate and a degree of freedom external to the substrate. A convenient model of the substrate is a patch of the Euclidean plane $\mathbb{E}^2$; the deviation might be anything—say, depth or color. Substrate and deviation are not necessarily commensurate. Formally, these are instances of trivial fiber bundles (Spivak 1975). A relief is a section of the bundle, a smooth assignment of a deviation to each point of the substrate.

Both local and global formal descriptions are often useful. In a local description, one typically differentiates between orders of approximation (see the sidebar titled Orders of Approximation and **Figure 1**). The first order (tangent plane, spatial attitude) was formally well understood in the seventeenth century (Barrow 1670). It is conventionally described by way of two angles, generally referred to as the slant and the tilt. In psychophysics, this description became understood

ORDERS OF APPROXIMATION

The formal description of pictorial relief involves a fiber bundle $\mathbb{E}^2 \times \mathbb{A}^1$. (Below, we simply write $\mathbb{A}$ for $\mathbb{A}^1$.) Points in the Euclidean plane $\mathbb{E}^2$ are parameterized by Cartesian coordinates $\{x, y\}$. The origin will be some fiducial point of $\mathbb{E}^2$. A fiber at some point is parameterized through an arbitrary point $\mathcal{O} \in \mathbb{A}$ ($\mathbb{A}$, the affine line) and equally arbitrary unit point $\mathcal{U} \in \mathbb{A}$. Then, any point on the fiber is designated $\mathcal{O} + z(\mathcal{U} - \mathcal{O})$, where the coordinate z is a pure (dimensionless) scalar.

It remains open how to relate points on distinct fibers, for each has its own $\{\mathcal{U}, \mathcal{O}\}$. A relation is imposed by way of what is referred to as a gauge field that relates distinct fibers to one another (Koenderink & van Doorn 2012). A pair of parallel planar cross sections (mutual separation Ξ, say) suffices for our discussion. Such a gauge would be an idiosyncratic set for a given observer.

Then, a cross section of the bundle can be described by a function $z(x, y)$. The Taylor expansion (the constant, or zeroth other term, will vanish identically) of depth becomes

$$z(x, y) = (a_{10}x + a_{01}y) + \frac{1}{2}\left(a_{20}x^2 + 2a_{11}xy + a_{02}y^2\right) + \cdots = \mathrm{I} + \mathrm{II} + \cdots, \qquad 1.$$

that is, the first (I) plus the second (II) plus higher-order (. . .) terms.

The slant of the first order is $\sigma = \sqrt{a_{10}^2 + a_{01}^2}$. Its magnitude is proportional with Ξ; thus only ratios of slants are meaningful. The tilt is the angle $\tau = \arctan(a_{10}, a_{01}) \in [-\pi, +\pi)$; it is independent of Ξ. Slant and tilt specify the first order.

The second order is described by the Hessian, a symmetric tensor with coefficients $\{a_{20}, a_{11}, a_{02}\}$. The Casorati curvature $C = \sqrt{(a_{20}^2 + 2a_{11}^2 + a_{02}^2)/2}$ also depends upon Ξ, so only ratios are significant. The shape index s (Koenderink & van Doorn 1992) is defined in terms of the ratio of the (real) eigenvalues and is independent of Ξ. The orientation φ—that is, the direction of the eigenvector belonging to the largest eigenvalue—is likewise well defined.

These properties of relief can be observed through methods of experimental phenomenology. All this is fairly recent. The first geometrical responses were obtained only in the 1990s (Koenderink et al. 1992). Gauge transformations were identified only after the turn of the millennium (Koenderink et al. 2001, 2011; Koenderink & van Doorn 2012). Most empirical work concentrates on the zeroth order, which is perhaps odd, given that no observer can really give a number in response to the question "What is the depth at this point?" where one might expect a "magnitude estimate" in the sense of Stevens (1957).

in the 1980s (Stevens 1983). The second order (osculating quadric, shape) yields a number of distinct categories. These were identified by Alberti (1435) in the fifteenth century, although a full treatment had to wait until Gauss (1827) in the early nineteenth century. Differential geometry (Strubecker 1942, Hilbert & Cohn-Vossen 1952, Spivak 1975, Sachs 1990) yields the proper formal description.

Global description is commonly used for geographical landscapes, in terms of hills and dales (Maxwell 1870), or ridges and ruts (Rothe 1915). This obvious language of relief is used by sculptors (Hildebrand 1893, Rogers 1969).

1.2. Visual Awareness and the Psychogenesis of the Visual Field

Visual awareness occurs when the eyes are opened (Koenderink 2013). It is prior to consciousness, which is unique to social animals, such as primates, and it does not even require a self. Consciousness implies awareness but not vice versa (Koenderink 2015a).

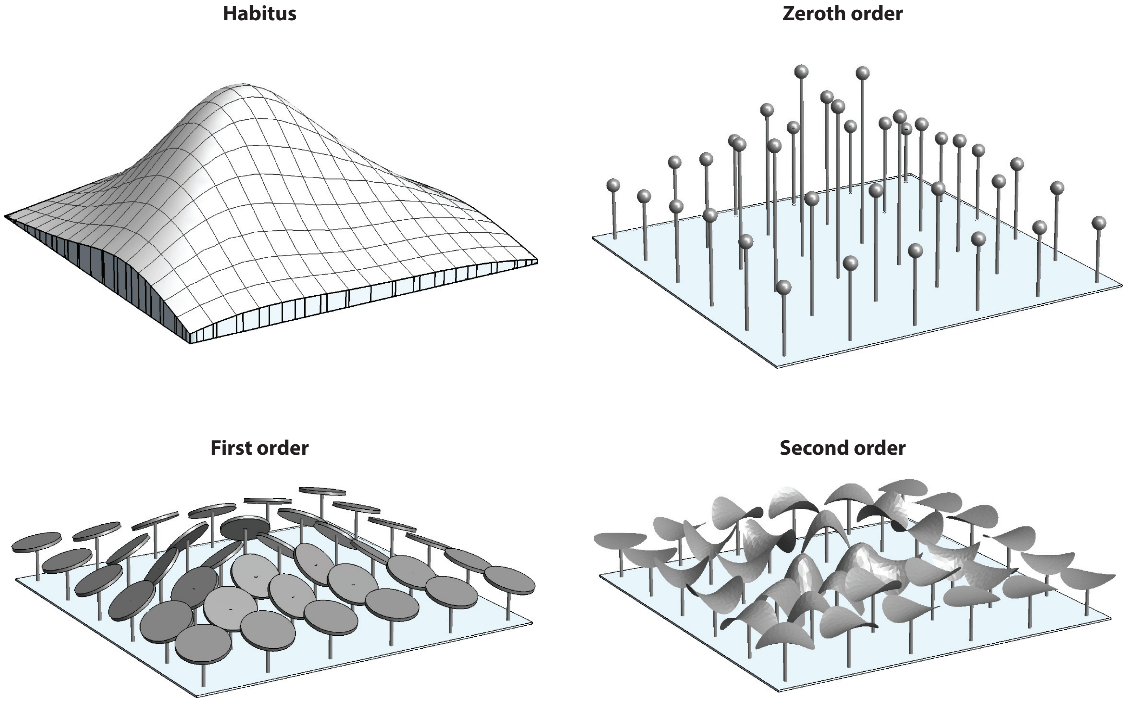

Figure 1

The notion of orders of representation. (*Top left*) A relief; the depth dimension is vertical. (*Top right*) The zeroth order; essentially depth samples. (*Bottom left*) The first order; these are surface attitude samples, representing the gradient of the depth. (*Bottom right*) The second order; these are local quadric deviations from the first order. The human observer mainly represents the first order.

The layout of visual awareness is the visual field. Geometry roughly models it as a trivial fiber bundle. The base space is a patch of the Euclidean plane $\mathbb{E}^2$, and the fibers are various qualities; think of tone, color, depth, etc. The visual field also contains values, and, indeed, any aspect of awareness is an emphatic value (Langer 1953, Koenderink et al. 2011, Koenderink 2014a).

For the sake of simplicity, consider the fiber bundle $\mathbb{E}^2 \times \mathbb{A}$ as a model. The base space $\mathbb{E}^2$ is the Euclidean plane, and the fiber $\mathbb{A}$ is the affine line.

Consider a tentative model of the psychogenesis of visual awareness (Koenderink 2015a). It formally connects phenomenology with optical structure, in a loose sense.

Affine line $\mathbb{A}$: geometrical entity that has no origin or unit of length; only ratios of lengths of intervals are invariant

Psychogenesis: a preconscious mental process that uses any of many alternative ways to come up with visual experiences

Awareness is a so-called brute fact, in that it cannot be explained scientifically. Brain science and cognitive science deal with objective facts; there can be no causal link between phenomenology and science. A psychophysical bridging proposal formalizes a kind of wormhole between ontologically disparate realms (Teller 1984). It is necessarily a heuristic, because it is neither objectively right nor wrong.

Schrödinger (1958) proposed such a heuristic. Any life-form actively pokes its environment in attempts to extend its dominion, and such poking poses questions to nature (Richards & Bobick 1988). Meeting resistance to poking is receiving answers. Schrödinger suggests that an unexpected answer comes with a "spark of awareness" (Koenderink 2015a, p. 123).

This idea implies perception as controlled hallucination (Clowes 1971), or analysis by synthesis (Yuille & Kersten 2006), as an apt heuristic model of psychogenesis (Hoffman et al. 2015, Koenderink 2015a). Controlled hallucination avoids inverse optics (Pizlo 2001), which—apart from being impossible—cannot give rise to meaning. Computation serves to transform structure into structure; thus meaning needs to be imposed. This is Vico's (1725) *verum factum*. Indeed, a human, a cat, a dog, and a dove are aware of very different meanings when confronted with a lantern pole, as implied by their behavior.

The psychogenesis of visual awareness implements questioning by shifting depth values along the fibers like beads on the strings of an abacus and judging the fit. Such "controlled hallucination" is a generic feature of depth relief, necessary to resolve the ambiguity inherent in all pictorial structure. (See also Adelson & Pentland 1996.) This important issue reappears frequently in the discussion.

Gestalts: generic templates of the optical user interface; occur automatically

Veridical vision: vision that implies an "all-seeing eye" view of the environment

1.3. Creative Imagery

Psychogenesis is a form of constrained imagery (Koenderink 2011). The constraint holds creative imagery against the contents of the visual front end. Imagery need not account for all front-end structure; in fact, ignoring the larger part of it tends to be advantageous. Psychogenesis selects substructures that it (tentatively) identifies as cues.

The front end is a volatile buffer that is continually overwritten by sampled, filtered, and formatted optical structure. The contents are due to the current fixation and thus codetermined by psychogenesis. Front-end activity is simply physical structure, and semantics need to be imposed. Psychogenesis wields so-called seek images and tends to apply template structures in order to do that. It has to come up with the best case it can make (hopefully beyond reasonable doubt) within a single moment.

Seek images are well documented in ethology (von Uexküll 1909). Humans are not all that different from their animal kin (Vallortigara et al. 1990, Spelke & Lee 2012). If one is dying for a beer, one is likely to spot bars everywhere; a very strong seek image dominates one's psychogenesis.

Template structures are elements of the optical user interface. Humans certainly hold many templates of the Gestalt type in common with animals (Spelke & Lee 2012). Vertebrate behavior (e.g., that of birds and fishes) is often largely controlled by releasers and imprints (von Uexküll 1909, Vallortigara et al. 1990). This is less obvious in primates, but even here, visual awareness is populated with optical user interface elements. Human user interfaces distinguish humans from cattle but are interfaces no less (Hoffman et al. 2015). User interfaces are idiosyncratic and shield the user from unnecessary complexity. They in no way promote veridical vision (Koenderink 2014b); indeed, veridicality can hardly be expected to promote biological fitness.

1.4. Pictorial Space and Pictorial Relief

This review focuses on the structure of pictorial space, especially pictorial reliefs.

Pictorial perception renders experimental phenomenology conceptually a reasonable undertaking because pictorial space is a mental entity without physical counterpart, so the topic of veridicality of perception does not come up. Methodologically, pictures are easy to present and make and change on the fly using computer methods. This turns experimental phenomenology into a practical undertaking. One studies purely mental entities using operations in two-dimensional (2D) assemblies of pigments—for that is what pictures are.

This review focuses on pictorial space to illustrate various properties of the geometry of visual awareness. (See the sidebar titled Looking Into Pictures.) In pictorial perception, many classical depth cues are ineffective with respect to pictorial content—for instance, binocular disparity, optical flow, accommodation effort, and optical blur. *Ineffective* is a misnomer, for these cues are very effective in revealing the flatness of the picture plane. They involve the picture as a physical object and work against the awareness of pictorial space. The instruments (popular in the past) to

LOOKING INTO PICTURES

One may look at a picture, or one may look into a picture. Sophisticated observers enjoy doing both simultaneously or flip voluntarily. Gibson conceived of pictures as views out of windows (Gibson 1950, Topper 1983). Then pictorial perception is a response to static, but otherwise generic, optical data. That is exactly what pictures are not—remember Maurice Denis's dictum. Paintings are flattish physical objects; they are all but looks out of a window. Picasso used to hang his paintings out of kilter so as to irritate visitors enough to notice them. Most painting techniques stress the picture plane. One looks into a painting and enters—as through a magic mirror—a miraculous space, fully unconnected to the space one moves in. Neither the eye nor the picture plane are in pictorial space. It is a mental entity without counterpart in the physical world. As Leonardo da Vinci (1540) noted, one may enter it by staring at a dirty wall.

view pictures to best advantage, like the zograscope, the verant, and so forth, use optical tricks to minimize their influence (von Rohr 1904, Zeiss 1907, Koenderink et al. 1994, Wijntjes 2017).

Pictorially important cues are conventionally known as "pictorial cues." Familiar instances are shading and occlusion.

A distinction between the physical/physiological and pictorial cues is that the latter allow rather ambiguous inferences. This is an important topic that accounts for many properties of pictorial awareness.

Consider a specific example, Shape From Shading (SFS) (van Doorn et al. 2011), which is due to a uniform, unidirectional beam of light falling on a surface composed of a uniform material. When does this lead to a uniform patch of radiance in the field of view? From elementary physics: only when the surface is planar. What about the location and spatial attitude of such a plane? Again, from elementary physics: completely arbitrary! Psychogenesis can check planarity but has to hallucinate overall depth and spatial attitude (Belhumeur et al. 1999; Koenderink et al. 2012a, 2013). This puts strong constraints on the geometry of pictorial space. It is a homogeneous space in which planarity conserving transformations play the role of congruences or similarities. Such transformations can be freely assigned (Koenderink & van Doorn 2012). They are referred to as mental eye movements (Equation 3; **Figure 2**).

Psychogenesis identifies planar patches modulo arbitrary eye mental movements. Patches often appear frontoparallel, even with a painting viewed obliquely (Koenderink et al. 2004). Psychogenesis controls both involuntary fixations (physical movements) and mental movements.

The group of mental eye movements was discovered only in 2001 (Koenderink et al. 2001). Its consequences are empirically borne out in quantitative detail, and much that used to be discarded as idiosyncratic observer differences suddenly turned into relevant data when considered modulo mental eye movements. Insignificant correlations suddenly turned into coefficients of variation (R^2s) over 0.99. The discovery changed the field decisively.

Mental eye movements: mental processes that empirically play a major role in the measurement of pictorial relief

2. OPERATIONAL DEFINITIONS OF PICTORIAL RELIEF

What is measurement? All measurement is comparison with some standard (Bridgman 1927). Consider length. An object is a yard high if one can set both the object and a yardstick on the floor, and one notices that their tops level up. If one has a graduated ruler, one similarly notices the coincidence of the top of the object with a mark on the ruler. Eddington (1928) spoke of "pointer readings." All it requires is the recognition of coincidence, not of quality or value. Classical physics developed this into a success story that lets one quantify all kinds of qualities. Even premodern

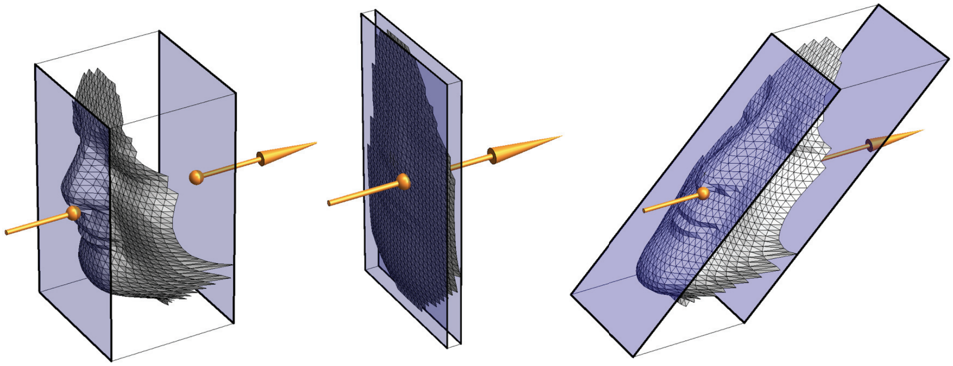

Figure 2

A depth slice subject to various mental movements. Compare the center and right images to the one on the left. The center image shows a strong depth scaling—a contraction—whereas the image on the right shows an isotropic rotation, which appears as a shear to the Euclidean eye. Transformations like these are routinely encountered when comparing observers in similar tasks. The intuitive picture of the two parallel panes is commonly attributed to Adolf von Hildebrand (1893). It is a geometrical model of a simple gauge transformation (Koenderink & van Doorn 2012).

societies developed ways to measure length, area, volume, and weight, all based on the recognition of coincidences.

How does this work out for pictorial qualities? Here, the object to be measured is in the mind, so how does one place the standard? Evidently, the standard needs to be placed in pictorial space too. There are numerous applications of this basic idea.

For instance, apparent surface slant is seen in pictures but can also be drawn in pictures. Artists do that all the time. They draw standard objects with the intention of letting one see the intended quality—in this case, surface slant. A conventional standard object is a circle or a planar disk. The draftsman introduces it everywhere, as ripples on a lake, the waterlily leaves on the surface, the hem of a dress, the wrinkles in a sleeve, the section of a head through the eyes, and so forth. (All conventional basic sketching classes will cover this.)

The draftsman's method is easily used in reverse. It can be turned into a tool to measure surface attitude (Stevens & Brookes 1987, Koenderink et al. 1992, Koenderink 2011). One superimposes an ellipse on the picture plane and has the observer adjust its shape so as to "fit" the pictorial relief. The only problem might be that the gauge object has to be seen as both an alien object and as located on the relief. If the gauge object "floats," there can be no comparison. If the gauge object "attaches," the coincidence is obvious, and it feels like the circle sucks on the surface.

As a method of measuring pictorial relief (**Figure 3**), this yields consistent results (Koenderink et al. 1992). Problems may arise from adroit use, for proper implementation requires something akin to the painter's eye. Improvements aimed at increasing precision or efficiency often tend to backfire for such reasons. Being able to "see" is something that every artist had to learn but many scientists lack (Koenderink 2015c).

Does the introduction of a gauge object affect the perception of the picture? Yes, any measurement is necessarily a perturbation, as is understood in physics and just as true in experimental phenomenology—hence the need to vary the method of measurement. Failing to do so counts as a methodological error. This also applies to the sciences proper. For instance, in astronomy, distances are measured by a variety of methods. Attempts to bring these in sync are very important in constructing an ideal entity that might be called "the" distance.

Fitting objects: objects that cling to the pictorial relief

Gauge objects: physical objects used for comparison as a kind of common currency; domains of experience are claimed for science after suitable gauge objects are found

Floaters: term used by artists to indicate objects that fail to join the intended pictorial space; examples include fly specks and craquelure on the surface of old paintings

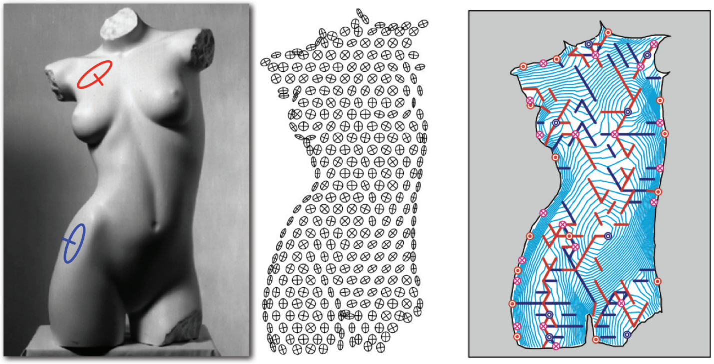

Figure 3

(*Left*) A stimulus, with two gauge figures (see the sidebar titled The Gauge Figure Method for Surface Attitude). Notice that the blue one fits and the red one does not; the latter even “floats” (fails to cling to the surface). (*Center*) Results from half an hour’s worth of settings of gauges by a human observer. (*Right*) A geometrical analysis of the relief. The light blue curves are equal-depth loci, the red lines are ridges of the relief, and the dark blue lines are ruts.

THE GAUGE FIGURE METHOD FOR SURFACE ATTITUDE

The basic idea is very simple: One superimposes an ellipse on the picture plane and judges its fit (the ellipse lies in the relief) by eye measure. If the fit is judged unsatisfactory, the shape and orientation of the ellipse are adjusted appropriately. One interprets the ellipse as a foreshortened circle, and then the foreshortening reveals the surface attitude. That would be an attitude measurement at one location.

To measure the relief, one needs to sample at multiple locations. Here, the problems soon accumulate. First, one can measure only one location at a time; otherwise, the accumulated gauge figures soon distort the relief. Second, the observer changes over time. This forces short, one-time sessions. Long periods or repeated sessions introduce problems of changing mindsets. As a consequence, one can sample only a finite number of samples. It is easy to sample dozens, and a dedicated, trained observer may sample hundreds. Thousands are forever out of reach.

To make the best use of, say, a hundred samples, one triangulates the patch and measures at the barycenters of the faces of the triangulation (of course, in random order). Such a set of samples is necessarily incongruent with a triangulated surface, so one has to fit the best approximation (in some optimal sense) to the data. Then, one judges whether the approximation statistically explains the samples. If so, one indeed has a measurement of the relief. If not, then there is no such a relief—at least, from the perspective of the method.

The formal check on internal consistency involves the identity of mixed second derivatives (i.e., $\partial a_{10}/\partial y \equiv \partial a_{01}/\partial x$), where a_{10}, a_{01} are the observed components of the spatial attitude.

In practice, one finds that there almost always is a relief. That shows that the glass bead game does not shift beads independently of one another. Psychogenesis fits template surfaces—at least, it would seem that way.

Various conceptual problems remain: Is there a surface in awareness, or preawareness, from which samples are taken? Or does the process of measurement create such a surface? And so forth. Such topics are addressed in Sections 2 and 3.

Gauge figure methods are desirable in experimental phenomenology because they implement the method of coincidence. A variety of different methods may complement or extend this; think of methods of reproduction or methods of absolute or differential judgments. The more methods, the better. The measure of a quality is necessarily indexed with the particular method. There is no guarantee that different methods address the same entity, like a length measured by a ruler may not be the length measured by echo location. A single word (e.g., *length*) often points at distinct entities. The convergence of measurements by different methods is a sign that one approaches some kind of so-called objectivity.

The German sculptor Adolf von Hildebrand (1847–1921) wrote a book *The Problem of Form* (Hildebrand 1893) in which he explains the structure of pictorial relief in detail. Like Michelangelo, his technique departed from the concept of relief. He would attack the block from the front and work his way into depth. He held that one also sees the finished work that way. By way of proof, he notices that observers have a hard time judging the depth of relief, the factor η of the similarity of the second kind. In that, he was right; observers often differ by as much as a factor of four when you measure the depths of their reliefs for the same stimulus. Hildebrand does not remark on the influence of the isotropic rotations, the vector $\mathbf{g}$ (see the sidebar titled The Geometry of Pictorial Space), although it is hard to doubt that he was aware of it. Idiosyncratic differences $\|\mathbf{g}\| > 1$ are common enough. (Notice that $\|\mathbf{g}\| = 1$ would imply a 45° slant in $\mathbb{E}^3$ but that such Euclidean angles are meaningless in $\mathbb{E}^2 \times \mathbb{A}$. The slant $\mathbf{g}$ is a non-Euclidean angle.)

Such huge effects are obviously of considerable interest. Other factors of interest concern the representation of the relief. Is it a distribution of depths, of surface attitudes, or of shapes (curvatures)? The gauge figure method addresses only attitude, ignoring depth or curvature, so

THE GEOMETRY OF PICTORIAL SPACE

Consider Cartesian coordinates $\{x, y, z\}$ for points $\mathcal{P} \in \mathbb{E}^2 \times \mathbb{A}$. Let $\{x, y\} \in \mathbb{E}^2$, and let $z(x, y) \in \mathbb{A}$ denote depth of relief. Define the unsigned distance subtended by points $\mathcal{P}_1, \mathcal{P}_2$ as $d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Notice that $d_{12} = 0$ in no way implies $\mathcal{P}_1 = \mathcal{P}_2$. When $z_1 \neq z_2$, one says that the points are parallel and defines $\delta_{12} = z_2 - z_1$ as a (signed) special distance. It is defined only in case the proper distance vanishes. This metric defines singly isotropic three-dimensional space. The affine line is taken as isotropic, with any points on it having zero proper mutual distance. The metric is left invariant by the group of transformations

$$\begin{aligned} x' &= h(x \cos\varphi - y \sin\varphi) + t_x \\ y' &= h(x \sin\varphi + y \cos\varphi) + t_y \\ z' &= g_x x + g_y y + \eta z + \tau, \end{aligned} \qquad 2.$$

when you set $h = \eta = 1$. Then φ parameterizes a rotation, and the vector $\mathbf{t} = \{t_x, t_y\}$ a translation in $\mathbb{E}^2$. Likewise, the vector $\mathbf{g} = \{g_x, g_y\}$ defines an isotropic rotation, and the scalar τ defines a depth translation. The parameter h, when different from 1, denotes a Euclidean similarity in $\mathbb{E}^2$ conventionally referred to as a similarity of the first kind. The parameter η, when different from 1, denotes a depth scaling in $\mathbb{A}$ conventionally referred to as a similarity of the second kind. The mental movements are the subgroup that conserve $\mathbb{E}^2$—that is, $x = x'$, $y = y'$. The shift τ is irrelevant since absolute depth is undefined. Thus the mental movements of psychogenesis are described by

$$z' = g_x x + g_y y + \eta z. \qquad 3.$$

The theory is empirically borne out in impressive quantitative detail, as shown below in Section 2. [Equation 3 is crucially needed for the analysis of pictorial relief (Koenderink et al. 2001, Koenderink & van Doorn 2012).]

what if one measures curvatures or depths directly? Does one such measurement predict results for a different one?

Other questions involve the influence of the pictorial structure or the viewing conditions. One often notices painters step back from their easel and squint or "screw up their eyes." What are they doing? Does it make a difference? Does a picture look different upside down or if seen in a mirror (which are methods of viewing suggested in books on how to paint)? There is no end to such questions. The field lies wide open for investigation.

2.1. Straightforward Pictorial Relief

The first truly geometrical measurement of a pictorial relief was published in the early 1990s. The stimulus was a scanned photograph of the Brancusi *Bird* at the Philadelphia Museum of Art (**Figure 4**). The triangulation involved 225 vertices, and a session took about a quarter of an hour. The result was a clean surface (Koenderink et al. 1992). It is an encouraging result, for consider how much relief data one might collect in a quarter of an hour using classical methods of psychophysics: There is simply no comparison.

There are obvious idiosyncratic differences, although the observers viewed the same picture. But a picture does not determine what was in front of the camera, so what is referred to as a beholder's share (Gombrich 1961) is expected. Observers were questioned after completion of the task. One observer interpreted a darkish spot on the photograph as a smudge, and another observer took it as shading due to vignetting in a cavity. Such effects are hardly surprising but cannot be predicted.

The gauge figure method for surface attitude (**Figure 3**) delivers remarkable results from comparatively minor effort. One obtains true geometrical structure as "response," instead of a collection of yes/no answers. The response is geometry; it allows formal geometrical analysis. One gains orders of magnitude in efficiency of observation this way compared to the results achieved from painstaking psychophysical experiments of the classical type. The price is that attitude settings are subjective; this is frankly experimental phenomenology. This disadvantage has the advantage that the result is meaningful. True psychophysical data are, by design, objective and

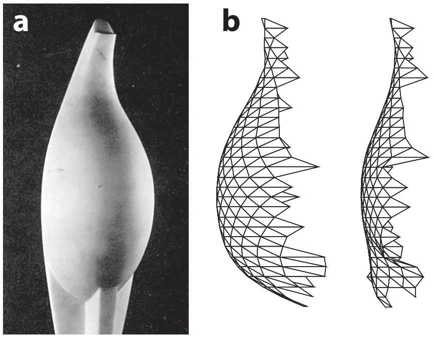

Figure 4

(*a*) The stimulus and (*b*) responses of two observers (depth increases toward the right). A response is a surface (cross section) in $\mathbb{E}^2 \times \mathbb{A}$. Notice the differences, both globally and locally. Some of the differences between the responses of the two observers can be removed by a gauge transformation and others cannot. Only the latter are relevant (Koenderink et al. 1992).

thus phenomenologically meaningless. Objectivity implies observations without an observer (von Foerster 1992) so fails to apply to phenomenology. There cannot be objective data on subjective meaning. The choice is either science or phenomenology.

If the choice is science, the topic of pictorial relief shrinks to almost nothing. So science is done soon, but, of course, it has other priorities. For phenomenology, the topic of pictorial relief is extensive. There are numerous directions to explore but no boundaries or final goals. It is like the visual arts in that respect.

Figure 4 shows an example. The two pictorial reliefs are obviously different; thus the results are indeed subjective, yet they are also similar, with the similarity apparently due to the structure of the image. In this case, the difference between the reliefs is mainly a depth scaling.

2.2. Issues of Veridicality

Questions about the veridicality of pictorial perception are often raised, since, apparently, large deviations from veridicality are common (Koenderink et al. 1996c). A direct check on veridicality is easy enough, but whether bothering makes sense is another matter. One method is to use a photograph of a physical object as the stimulus. Then one may immediately compare the response with a photograph of the same object seen from an orthogonal direction (Koenderink et al. 1996c, Todd et al. 1996), for the latter photograph explicitly depicts the depth encoded in the former (**Figure 5*a***).

Not surprisingly, the response is not identical to this ground truth. Yet the response is surprisingly close to the physical layout modulo a gauge transformation, with the deviations being mainly an isotropic rotation and scaling.

Another way to judge veridicality is to consider differences between responses by different observers for the same picture. Here, one typically finds differences that can be accounted for in terms of idiosyncratic mental eye movements. It is easy enough to test for that; one simply does a regression modulo arbitrary mental movements (Koenderink et al. 2001). Doing that, typical correlations between depth values at corresponding locations for different observers are in the 0.9 or higher range for regular photographs of simple objects (**Figure 5*b***).

The effect of mental movements is crucial. For the photograph of the Brancusi turtle (**Figure 6**), the straight depths correlation yielded $R^2 = 0.06$, whereas this became $R^2 = 0.95$ for

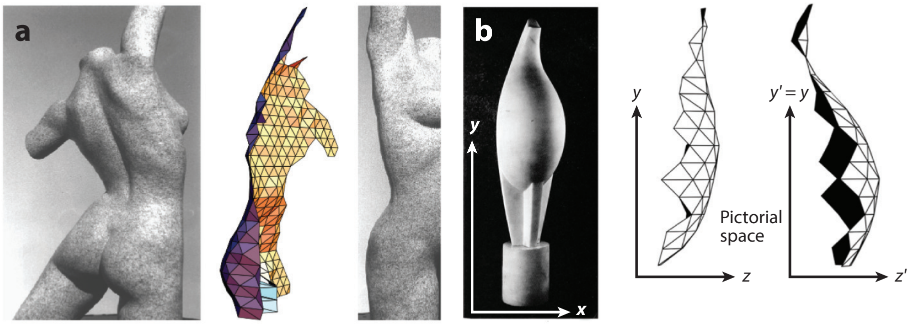

Figure 5

Two checks on veridicality. (*a*, *left*) The stimulus, (*center*) a relief (depth increases toward the right), and (*right*) the ground truth. (*b*, *left*) The stimulus, and (*right*) two reliefs by different observers (depth increases toward the left). The *x*, *y*, and *z* coordinates relate to Equation 3. Notice the apparent mutual rotations: The responses are almost identical modulo mental eye movements (Koenderink et al. 2001).

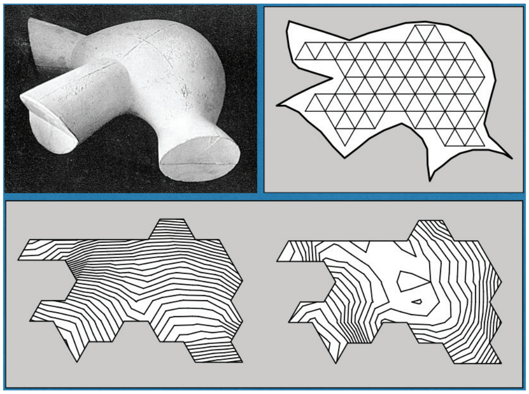

Figure 6

(*Upper left*) The stimulus and (*upper right*) its triangulation. (*Bottom*) Two reliefs for the same observer, resulting from two only slightly different tasks (Koenderink et al. 2001).

a regression modulo mental movements. Two reliefs for the same picture might be considered essentially uncorrelated when the effect of mental movements is ignored in the analysis. However, they may become essentially identical when the analysis takes proper account of the effects of mental movements (Koenderink et al. 2001). Failure to recognize mental eye movements renders reports uninteresting, because major features of the data are hidden when conventional methods are applied uncritically.

Synoptical views: binocular views with identical input to the two eyes; optical devices like the Zeiss synopter allow one to annul binocular disparity

2.3. Effect of Viewing Conditions

Every artist knows that viewing conditions are crucial. Whether one closes an eye when looking into a painting, for instance, makes a huge difference. Measurements of pictorial relief immediately reveal the importance of this insight.

Although the nature of the relief remains the same, the depth of relief is very different for monocular and binocular viewing (Koenderink et al. 1994). (See the sidebar titled Stereopsis.) The effect is even larger for synoptical viewing, a method that removes binocular disparity by optical means (**Figure 7**). These effects are readily described in quantitative detail via mental movements—in this case, through similarities of the second kind (Equation 3).

STEREOPSIS

Stereopsis literally means spatial (3D) vision. It originally applied to natural vision. The only exception was seeing as a newborn child and enjoying a chaotic 2D quilt of colored patches. This may be natural for newborns, but only Claude Monet is quoted as a heroic, mature example. Generic vision is spatial whether monocular or binocular, static or moving; even the creative imagination (no disparity!) is spatial.

This changed with behaviorism and its direct heir, cognitive science. Recent dictionaries define stereopsis as binocular, and monocular vision as flat. This would have been considered nonsense only a century ago (Emerson 1863, Claparède 1904, Ames 1925, Schlosberg 1941, Pollack 1955, Schwartz 1971, Higashiyama & Shimono 2012). In 2014, Vishwanath (2014) could write “Towards a New Theory of Stereopsis.”

Generic (spatial) vision is based on monocular stereopsis, although motion parallax and binocular disparity affect awareness too. Cases where a single cue determines awareness are laboratory artifacts.

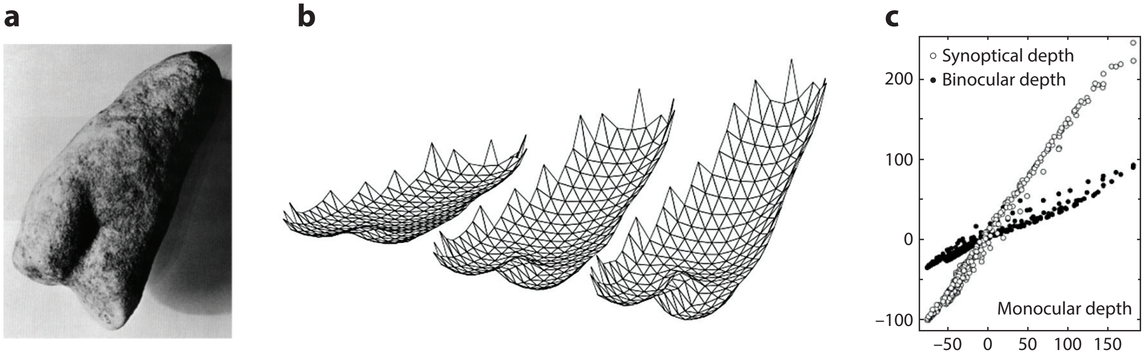

Figure 7

(*a*) The stimulus, (*b*) three reliefs (depth increases upward) for the same observer under different viewing conditions (from *left* to *right*: monocular, binocular, and synoptical viewing), and (*c*) a scatterplot of depths for the viewing conditions. The neat linear relations prove that the relation is an almost perfect mental eye movement, a similarity of the second kind (Koenderink et al. 1994).

Somewhat different results are obtained by reflecting images left/right or up/down (Cornelis et al. 2003, 2009). This is a method often used by artists to obtain a fresh view of a picture. In this case, the differences are again due to mental movements, but here, these are not similarities but isotropic rotations. The finding is perhaps intuitively reasonable. What is interesting is the magnitude of the effects, which is quite large.

It is often thought that so-called true relief must be due to binocular disparity or movement parallax (Rogers & Graham 1982). Yet the spatial resolution of the disparity relief is an order of magnitude worse than visual acuity (Tyler 1975). Indeed, monocular stereopsis often overrules binocular disparity (Cornilleau-Pérès et al. 2002, Koenderink 2015b). Relief is primarily due to the pictorial cues, although the physical/physiological cues all play a—not very well-understood—auxiliary role. This huge field of enquiry remains largely unexplored.

2.4. Effects of Pictorial Structure

It may be naively thought that only the geometrical structure of the picture is of relevance. This is not at all the case; for example, it makes a difference if one changes the illumination of a photographed object. Photographers and cinematographers make a profession out of that. Such effects are easy enough to study. One simply makes sure that the geometry of the camera with respect to the environment is unchanged and takes exposures in different illumination conditions.

One finds that the reliefs are remarkably invariant against illumination changes; thus there is a high degree of constancy. However, significant and systematic differences are easy to detect in the responses. They are conveniently studied through principal components or a factor analysis of the reliefs (Koenderink et al. 1996a).

Some of these changes may be explained through mental eye movements. However, one finds that shape changes that cannot be accounted for this way occur; thus the relief changes intrinsically. What roughly happens is that brightly illuminated parts tend to protrude, as if the light source "pulled" on the shape (Koenderink et al. 2012a, 1995). Such methods allow one to study the effects of shading per se, modulo the masking effect of the idiosyncratic mental eye movements. This opens up a novel field of enquiry.

Various parts of the relief react very differently to a change of illumination of the object from frontal to oblique. In frontal illuminations, many articulations of the form become almost hidden,

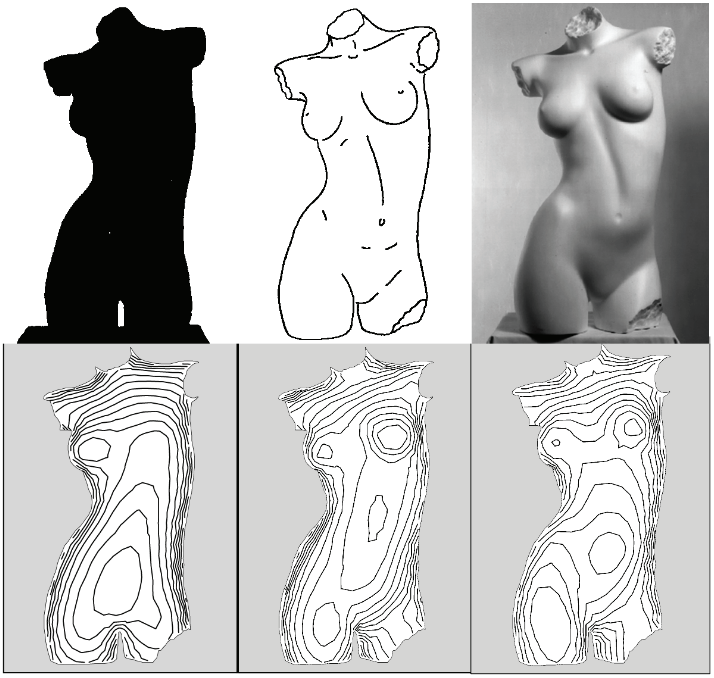

Figure 8

(*Top row*, *left*) A silhouette and (*middle*) a cartoon rendering based on (*right*) a photograph. Each column shows a stimulus on top and corresponding relief—all for the same observer CC—on the bottom (Koenderink et al. 1996b).

whereas an oblique illumination reveals shallow ridges or ruts that run across the direction of illumination. Psychogenesis restricts mental eye movements to limited regions of interest (Todd et al. 1996). This makes ecological sense, as it is rarely the case that global lighting applies over extended regions.

2.5. Silhouette and Cartoon Rendering

Another way to study the influence of pictorial content on relief is to keep the overall pictorial geometry constant but vary the articulation of internal detail. In the extreme case, one might even substitute a silhouette for the smoothly shaded object (Koenderink et al. 1996b). A slightly less drastic transformation would be a substitution of a cartoon drawing for the exact same outline as the smoothly shaded photograph (**Figure 8**).

Even in the case of the silhouette, one obtains an articulated relief. Whether the silhouette shows a ventral or dorsal view is fully ambiguous—if "shows" is the right term here—so psychogenesis needs to choose. Sometimes one encounters mixtures of ventral and dorsal views. A striking feature is that the curves of equal depth tend to run roughly parallel to the contour in the case of the silhouette, whereas they have a rather more complicated relation to the contour in the case of the fully shaded image.

Cartoon renderings yield results that are close to those obtained for fully shaded images, although the gauge figure is inside empty white spaces between the drawn lines. That observers inject idiosyncratic templates is evident from the fact that much detail is apparently "hallucinated." Mechanisms for such ability have been suggested [such as Biederman's geons (Biederman 1987,

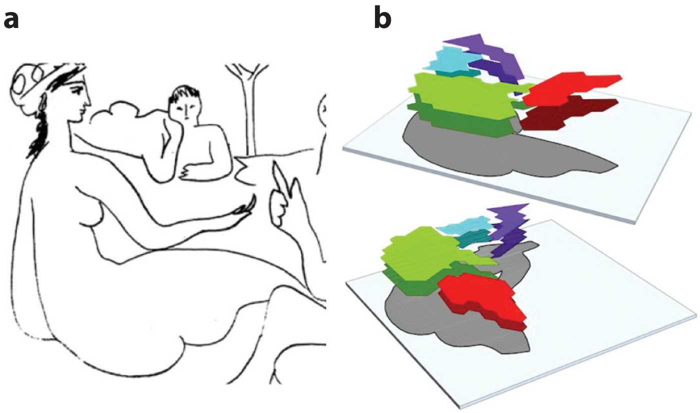

Figure 9

(*a*) A minimalistic Picasso drawing and (*b*) the parts from a cluster analysis on the reliefs of several observers. The maps of the gauge transformations applied by two of the observers with respect to the mean over all observers are clearly different. Notice that they differ per part (Koenderink et al. 2012b).

1988)]. Contours can be very suggestive of 3D shape, especially in drawing (Crane 1900, Kennedy & Silver 1974, Hoffman & Richards 1984, DeCarlo et al. 2003).

Although relief is certainly constrained by the structure of the optical stimulus, the nature of relief also reveals creative imagination. The blend of stimulus-determined and observer-determined contributions varies, and obviously, the expertise of the observer is important. An example of such so-called injected knowledge was found in reliefs for two observers in the case of a minimalistic Picasso drawing (Koenderink et al. 2012b). In the drawing, the spine is not indicated, except by the mutual "gripping" by the left and right contours of the thorax. The response of one observer has a clearly articulated spinal groove in the dorsal area, whereas the response of the other observer has a generic convexity, a kind of balloon-shaped interpolation, as one notices in the case of a mere silhouette. It implies that different observers are aware of different reliefs when confronted with the same picture. No doubt, if the spine is found in the relief, it is present in visual awareness, even though not present in the stimulus. Awareness is not a function of the stimulus alone.

Cartoon renderings, such as minimalistic line drawings by artists like Picasso, yield excellent stimuli to study the effect of the beholder's share over a number of observers (Koenderink et al. 2012b). One finds that observers treat the relief in patches that correspond to natural body parts as suggested by the rendering (**Figure 9**). For each individual patch, observers agree quite well when the effects of idiosyncratic mental eye movements (the transformations described in the sidebar titled The Geometry of Pictorial Space) are taken into account. However, these gauge transformations are apparently assigned on a per-part basis, and choices differ with each observer. One surmises that observers are aware of different poses when perusing the drawing (Cornelis et al. 2003, 2009; Koenderink et al. 2012b, 2015b).

3. EXPERIMENTAL PHENOMENOLOGY OF RELIEF

So far, the review has mentioned a variety of observations mainly based upon a single method, that of surface attitude samples by way of elliptical gauge figures. This sufficed well enough to illustrate quite a number of important generic phenomena. However, it leaves numerous open questions—some of a methodological nature, others conceptual.

Lipschitz condition: condition posed as $|z_{12}| \leq M d_{12}^{\alpha}$ with $0 < \alpha \leq 1$; this is also known as the Hölder condition with Hölder exponent α

Consistency checks: the unique way of rendering phenomenological observations objective; they yield a bridge between experimental phenomenology and science

3.1. Some Methodological Issues

Whatever the method of measurement, one at most obtains a finite set of samples, supposedly relating to some ideal [in the sense of Meinong's (1899) "subsistent" objects] surface. Whether such an integral image actually exists in the mind is a conceptual issue addressed later.

It is a methodological issue whether a set of samples may be formally accepted as representing a surface. Various checks may be used, as *surface* has various definitions.

In the zeroth order, the simplest is perhaps a Lipschitz condition of continuity. Two samples at nearby locations should clearly have similar depth values, which makes sense only once a gauge has been fixed. A Lipschitz condition (Lipschitz 1864) states that $|z_{12}| \leq M d_{12}$, where d_{12} is the proper distance between the samples, $|z_{12}|$ is the absolute depth difference in the gauge, and M is a Lipschitz slope.

For higher orders, differential geometry suggests additional conditions. For instance—as noted earlier—the field of attitude samples should be a gradient field, thus have vanishing curl. The implementation of such checks is a formal, technical matter of some complexity (Koenderink et al. 1992). It also requires much additional empirical work to estimate the natural spread in the observations. In principle, one has to use repeated settings, because there is no notion of a ground truth. However, any repeat will in principle change the observer.

Additional checks become available when samples of different orders have been observed (Koenderink et al. 1996c, 2011). For instance, the first order is the derivative of the zeroth order, whereas the zeroth order is the integral of the first order. Integral surfaces for first-order samples were shown in previous figures. They make sense only if the field of samples is integrable (in this case, a gradient field) and only up to arbitrary constants of integration (in this case, absolute depth). Similar considerations apply to higher orders; they are indicated by the differential geometry.

Although this may perhaps appear merely technical, it is of great importance, because it implies that observations of experimental phenomenology are subject to precise, formal consistency checks. This deserves to be a major goal of vision science, although admittedly—and perhaps regrettably—not much has been achieved to date. The results discussed above indicate that such a goal is not at all elusive.

An important conceptual issue, not covered in this review, is which structures determine the settings used in obtaining a sample. By using presentations confined to small regions of interest, we immediately find that it is obvious that the task is impossible unless a large part of the picture is simultaneously displayed at all times; thus even local properties of the relief are globally determined (Erens et al. 1993). Eye movement studies reveal that observers always look at the gauge figures when doing their settings, though they scan the whole stimulus. Here lie many perplexing problems that need to be addressed. Despite the fundamental importance of the area of interest used in the gauge figure setting task, there appears to be no literature on the topic.

With respect to eye movements, a study by Enright (1987) is of interest. This author showed that vergence movements accompany the view of line drawings suggesting 3D shapes. The magnitude of these movements is rather idiosyncratic. This may be taken as a physiological sign of the activity of the psychogenesis of relief. It is not known whether a simultaneous awareness of the picture surface would suppress such movements. In his Pandora's Box apparatus, Gregory (1968) inverts this and attempts to measure absolute depth through matching with the vergence for a superimposed mark. Here, one has an interesting interaction between physical and pictorial space. Very little is known about the phenomenology of such interactions. More research is sorely needed, for although the method is used in applications (Chevrier & Delorme 1983, Bülthoff & Mallot 1988), it is hardly understood conceptually.

3.2. Conceptual Issues

One major aim of research on pictorial relief should be to develop and investigate as many methods of measurement as humanly possible. Here, humanly possible implies that observers should experience the task as "natural" and that a useful data volume can be sampled within the time span of a roughly constant mental set. Both objectives are crucial.

A task feels natural if it does not require reflective thought and does not frustrate sensorimotor interaction. For instance, asking an observer to call out an estimated depth [a number on a "magnitude estimation" scale (Stevens 1957)] for a location indicated on a relief (say, by placing a red dot on a monochrome photograph) does not feel natural because it requires reflective thought—for example, "Where is the origin?" or "What is the unit?" Many observers would object to the task.

Practical problems tend to be due to adroit interface design. Interface design is something of an art. When a task takes great dedication due to an awkward interface, this is likely to affect the results.

Other methodological issues involve the perturbation of the stimulus by the gauge object. A common example is to show all gauge figures simultaneously, like illustrated in **Figure 3**. In the example, the relief can even be seen in the absence of the stimulus, evidently not acceptable at all.

The time constraint is crucial because observers tend to change their mental set—in this case, the gauge. Running the same experiment repeatedly on the same observer always results in observations that differ by some mental eye movement. We find that dedicated observers may maintain their gauge for half an hour, or (at most) an hour. Naive observers, hired for the occasion, may be okay for a quarter of an hour or so. But having a cup of coffee or visiting the bathroom, perhaps even having a distracting thought, should imply termination of a session.

This imposes some important constraints on what can be achieved. It still leaves a huge area of endeavor, though most important conceptual problems in this area have hardly been touched upon but are easily within reach of existing methods.

3.2.1. The zeroth order. Magnitude estimation yields an impossible task. But given a pair of locations, most observers find it natural to judge which one is nearer in depth, especially when the locations are not too far apart. It is indeed possible to obtain a reasonably sized sample this way. For larger sets of points, one has to subsample. For instance, the number of pairs grows much slower when considering only neighboring points. One obtains results that are very similar to what is obtained with the elliptical gauge figure. In a quantitative analysis, one finds that the attitude samples suffice to account for the pairwise judgments, but not vice versa (Todd & Reichel 1989, Koenderink & van Doorn 1995, Reichel et al. 1995, Koenderink et al. 1996c). This appears to indicate that the mental representation is more likely to be surface attitude than depth per se. However, the available evidence is rather sparse.

It is conceptually much more interesting to compare points that are far apart than points that are close. Here, the subsampling may involve a smallish number of fiducial points (say, $M \ll N$), with each compared with a larger number (say N) of sample points. This indeed yields remarkable results.

One finds that only two points on a single slope of the relief are immediately comparable (Koenderink & van Doorn 1995). This suggests that the integral surfaces as illustrated above are not available for depth comparison in visual awareness. It might appear strange that this is the case, because the integral surfaces were obtained from straight responses—that is, from immediate visual awareness. One could speculate that the mental representation in awareness is not an integral surface but perhaps more like a gradient field. Integral surfaces are then objects that exist only in the analysis, rather than the awareness per se.

3.2.2. The first order. The attitude samples from the elliptical gauge figure involve two degrees of freedom—say, slant and tilt. Are these degrees of freedom individually available? This is easy enough to check; one lets the observer indicate either the orientation of constant depth or the direction of steepest depth increase. Observers find those natural tasks and the results are fully accounted for by the gauge figure samples.

Do the observers' responses shown in **Figure 4** have different total depth range? At first blush, it might seem so. However, the difference is apparent only if one reckons modulo mental eye movements. To answer the question, one needs to find the number of just-noticeable depth differences contained in the total depth range (Koenderink et al. 1992, 1996c). One finds that dedicated observers resolve about a hundred depth levels. For naive observers, this can be much less, all the way to zero. In the latter case, it may be said that such observers have no pictorial depth. So far, only sparse results are available though. The reason is that estimating the just-noticeable differences takes considerable effort (often times, repeated sessions).

3.2.3. The second order. As discussed above, there are reasons to believe that a mental representation is not so much in terms of depth (the zeroth order) but more likely surface attitude (that is, the first order). But what about higher orders—say, the second order? It is not entirely strange to guess that a mental representation might exist in terms of shape (or curvature) instead of spatial attitude (Stevens 1990). After all, curvature is invariant with respect to mental eye movements, whereas spatial attitude depends upon the gauge. To address such questions, one needs to observe shape samples.

Such an investigation is a major undertaking because very little is known concerning the perception of surface curvature (Erens et al. 1993, Johnston & Passmore 1994, De Haan et al. 1995, Koenderink et al. 2012a, Koenderink 2015d). Indeed, the formal understanding of surface shape is largely lacking in experimental psychology. Only a few studies are available (Koenderink et al. 2014, 2015a). A preliminary analysis suggests that the primary representation is in terms of surface attitude, rather than local shape. This is perhaps slightly counterintuitive, because shape is invariant, whereas attitudes are not. However, it might be that a shape-based representation would not be a curvature field but rather a qualitative, semiquantitative representation in terms of more global entities (see Section 3.2.4 below).

3.2.4. Global properties of relief. Global properties that are readily available for investigation are the geographical features. One thinks of the following:

- peaks, pits, and passes (or saddles) as special points
- mountain crests (ridges) and river courses (ruts) as special curves
- hills and dales [Maxwell's (1870) natural districts] as special areas

Notice that these are defined only in some given gauge. These features are easily found from empirically determined reliefs (e.g., from surface attitude samples), but are they immediately available from the mental representation?

So far, there appears to be no direct empirical data on this issue. It is clear that observers have a notion of the landscape-like features, but it is hard to let them trace these. Most observers will readily indicate the special points (Phillips et al. 2003) but shrink from the curves and areas, using the excuse that they are not able to draw. Methods to get around this tend to be time consuming. For instance, observers readily indicate whether two points are in the same natural district (Koenderink et al. 2015b), but finding a full segmentation this way involves a huge number of trials.

However, there is quite a bit of indirect evidence that the geographical entities play a role. In studies of interobserver differences, one finds very clear signs of idiosyncratic gauges applied to natural districts. Thus psychogenesis apparently segments the relief into natural districts and applies distinct mental eye movements to each (Koenderink et al. 2012b, 2015b).

Participants find drawing somewhat doable in the 1D (but not the 2D) case. If one draws a straight line over the picture plane, many are willing and able to draw a cross section in depth. Using computer graphics methods, this is easily turned into a viable method of obtaining global impressions of relief (Koenderink et al. 2001), and the method is very fast. One uses a regular triangular grid and considers one straight grid line after the other, in random order. This yields a significant oversampling, thus enabling a valuable check on the consistency of the data. As expected, participants tend to apply different gauges for about any trial. Thanks to the oversampling, one may derive an integral relief surface modulo arbitrary gauges on all lines though. This renders the analysis somewhat complicated, but the results are very useful. The reliefs presented in **Figure 6** were obtained in this manner.

One finds that the reliefs found in this way are very similar to those obtained by way of local sampling methods, such as the gauge figure method for surface attitude or the two-point comparison task. This indicates that there is indeed some kind of common substrate in visual awareness that emerges with all these methods.

4. CONCLUDING REMARKS

This rather short review necessarily presents a limited perspective on pictorial relief. In a full book format, one would almost certainly have a fair number of chapters presenting similar tunnel visions, with only sparse mutual interconnections. This indicates that the topic of pictorial relief is still somewhat of a quilt, instead of a coherent, well-integrated area of understanding. There are a number of quite distinct reasons for this, and we summarily discuss them in this final section.

Presentation and viewing modes make a huge difference. In this review, we stressed the stimulus to be a picture—that is, a physical object that exists quite independent of pictorial content. Call this picture mode.

In contradistinction, one may discard the existence of a picture altogether and assume Gibson's notion of a picture as the view out of a window. Call this aperture mode.

Something in between is a presentation that attempts to remove all cues to the existence of a picture plane yet is evidently a display instead of an aperture. Gregory, in his Pandora's Box method, used a back-illuminated translucent plate to remove all textural cues of the picture plane, for instance. Call this display mode. It is common in computer graphics presentations.

Picture mode is stressed through textural cues (e.g., the canvas in a painting), the picture frame, and limited angular size. This implies short viewing distance and an illuminated, articulate environment. Aperture mode reverses all that. As a result, there is no such thing as a so-called locatedness with respect to the eye for isolated pictorial objects in aperture mode, whereas there is an imprecise relation between pictorial and physical space in picture mode. This is evident from occasional remarks that a pictorial actor breaks the picture plane (e.g., sticks a foot or hand out of the picture frame). In display mode, it is of importance whether the observer has a sense of the environment. Even in a darkened room, the observer is usually—perhaps dimly—aware of the presence and position of the display. The pictorial content will be roughly felt to "be" near the display, even though pictorial space is distinct from physical space. It is not always possible to glean the relevant setting from literature data.

Genericity: the formal way to exclude special, singular cases; cases that occur with probability zero (though still possible) are excluded

In one experiment, the gauge figure method was applied to a real physical scene using a view-through-the-window view. The gauge figure was projected from the eye location on the physical object. There appeared to be no essential differences between pictorial relief and view-through-the-window relief (Koenderink et al. 1995).

In another study, point correspondences were obtained for 3D physical objects in aperture view. This involved a case in which no corresponding pictorial data were available (Koenderink et al. 1997). It would appear that there are many instances in which the pictorial and aperture presentation would yield similar results. Differences due to physiological cues are to be expected though.

Choice of pictorial elements makes a huge difference. The examples presented in this review involve either photographs of actual (physical) scenes or hand-drawn line pictures. There is a huge amount of literature in which the pictorial elements were chosen to be the major or, ideally, the only cues to relief. This usually implies display or aperture mode. In some cases, the choice also dictates the angular field of view—for instance, when linear perspective is stressed or when textural cues strongly dominate (Todd et al. 2005, Saunders & Backus 2006, Todd et al. 2007).

When the bouquet of cues is intentionally restricted, one forces psychogenesis in some specific mode. This may be advantageous from a reductionist perspective, but it has the disadvantage that it may be quite hard to generalize from the results. In a natural image, one invariably encounters a complicated blend of cues that all add their voice to the overall choir, so to speak (Wagemans et al. 2010, van Doorn et al. 2011, Koenderink et al. 2016b). Few of these cues would work in isolation, but if they do work, they would do so in a manner quite distinct from the generic case.

For instance, if one isolates the shading cue, pictorial relief becomes extremely ambiguous or even vanishes (Erens et al. 1993). One is aware of some so-called misty space. If one adds a contour, the shading gives rise to very different surfaces, dependent on the shape of the contour (Wagemans et al. 2010, Koenderink et al. 2012a). Such observations hardly reveal the importance of shading in natural images though. To address that, one needs to subtly vary the shading about some natural set point and correlate with the resulting changes of the relief. Such a technique can be applied to almost any cue but is, perhaps regrettably, rarely used.

Choice of the geometry makes a huge difference. Naively, it appears only natural to use as simple a geometry as possible in an attempt to understand the psychogenesis of pictorial relief from a formal perspective. Thus a large part of the literature concentrates on planes, cylinders, and quadrics (typically spheres or ellipsoids), or combinations of these (e.g., two planar patches connected by a hinge to a dihedral edge). The problem with such choices is that very simple surfaces imply ambiguous stimuli. Psychogenesis is forced to resolve the ambiguity and thus be in an unnatural mode, making generalization from the results difficult. The mathematician's notion of geometrical simplicity as genericity—that is, invariance with respect to small perturbations—applies immediately to vision. Of course, one needs ideas concerning the type of perturbations that are relevant to the problem.

For example, an isolated planar patch has no well-defined spatial orientation in pictorial space. It may appear in almost any attitude, with the default being frontoparallel (Koenderink et al. 2004, 2016a). Thus the perceived attitude will be very volatile and tends to be determined by minor factors that are not always easy to foresee.

These problems are even more pressing because most of the literature fails to take into account the—often major—effects of gauge transformations due to mental eye movements.

The upshot is that large pieces of the literature essentially define islands of knowledge that remain unrelated. As noted above, we can hardly cover that in this short review.

SUMMARY POINTS

Empirically, observers agree quite well on pictorial relief, albeit modulo mental eye movements. These are equivalences in view of the unavoidable ambiguities of pictures. Thus idiosyncratic elements are a necessary element of visual awareness. Key elements to keep in mind are the following:

1. Pictures are by their very nature ambiguous. Many visual presentations are equivalent in that they account for the optical structure equally well. These equivalences necessitate an important beholder's share (perspective of phenomenology).
2. Visual awareness is an optical user interface (the perspective of ethology). The inverse optics of machine vision does not apply.
3. Imagery starts from seek images and attempts to account for optical structure in the visual front end (perspective of biological mechanism).

FUTURE ISSUES

1. What the essential representation of pictorial relief is like is not known. One guesses it represents local surface attitudes modulo mental eye movements. At present, sufficient observations do not exist to allow a decision.
2. Psychogenesis seems to apply different gauge fields in different regions of interest. It may well be that this would be task dependent. After all, an image has many potential meanings, and observers may become aware of different structure once cued to it.
3. The split-and-merge operations that define regions of interest are ill-understood. Such segmentation is an essential requirement for the possible application of inverse optics algorithms, for the necessary preconditions at best apply to limited regions.
4. Psychogenesis certainly leaves geometrical incoherences in its presentations. This should be familiar enough from the popular "impossible figures." Their nature is barely understood.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The work was supported by the DFG Collaborative Research Center SFB TRR 135, headed by Karl Gegenfurtner (Giessen University, Germany), and by the program by the Flemish Government (METH/14/02), awarded to Johan Wagemans. Jan J. Koenderink was supported by the Alexander von Humboldt Foundation.

LITERATURE CITED

Adelson EH, Pentland AP. 1996. The perception of shading and reflectance. In *Perception as Bayesian Inference*, ed. D Knill, W Richards, pp. 409–23. New York: Cambridge Univ. Press

Alberti LB. 1972 (1435). *Della Pittura* [On Painting]. London: Penguin Class.

Ames A Jr. 1925. Depth in pictorial art. *Art Bull*. 8:5–24

Barrow I. 1916 (1670). *The Geometrical Lectures of Isaac Barrow*, transl. JM Child. Chicago: Open Court Publ. Co.

Belhumeur PN, Kriegman DJ, Yuille AL. 1999. The bas-relief ambiguity. *Int. J. Comput. Vis.* 35:33–44

Biederman I. 1987. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94:115–47

Biederman I. 1988. Surface versus edge-base determinants of visual recognition. *Cogn. Psychol.* 20:38–64

Bridgman PW. 1927. *The Logic of Modern Physics*. New York: MacMillan Co.

Bülthoff HH, Mallot HA. 1988. Integration of depth modules: stereo and shading. *J. Opt. Soc. Am. A* 5:1749–58

Chevrier J, Delorme A. 1983. Depth perception in Pandora's box and size illusion: evolution with age. *Perception* 12:177–85

Claparède E. 1904. Stereoscopie monoculaire paradoxale. *Ann. Ocul.* 132:465

Clowes MB. 1971. On seeing things. *Artif. Intel.* 2:79–112

Cornelis EVK, van Doorn AJ, de Ridder H. 2003. Mirror-reflecting a picture of an object: What happens to the shape percept? *Percept. Psychophys.* 65:1110–25

Cornelis EVK, van Doorn AJ, Wagemans J. 2009. The effects of mirror reflections and planar rotations of pictures on the shape percept of the depicted object. *Perception* 38:1439–66

Cornilleau-Pérès V, Wexler M, Droulez J, Marin E, Miège C, Bourdoncle B. 2002. Visual perception of planar orientation: dominance of static depth cues over motion cues. *Vis. Res.* 42:1403–12

Crane W. 1900. *Line and Form*. London: G. Bell & Sons

da Vinci L. 1540. *Libro di Pittura*. Vatican City: Bibl. Apostol. Vaticana

De Haan E, Erens RGF, Noest AJ. 1995. Shape from shaded random surfaces. *Vis. Res.* 35:2985–3001

DeCarlo D, Finkelstein A, Rusinkiewicz S, Santella A. 2003. Suggestive contours for conveying shape. *ACM Trans. Graph.* 22:848–55

Denis M. 1890. Définition du neo-traditionnisme. *Art Crit*. 65:540–43

Eddington AS. 1928. *The Nature of the Physical World*. Cambridge, UK: Univ. Press

Emerson E. 1863. On the perception of relief. *Br. J. Photogr.* 10:10–11

Enright JT. 1987. Perspective vergence: oculomotor responses to line drawings. *Vis. Res.* 27:1513–26

Erens RGF, Kappers AML, Koenderink JJ. 1993. Perception of local shape from shading. *Percept. Psychophys.* 54:145–56

Gauss CF. 1827. Disquisitiones generales circa superficies curvas. *Comment. Soc. R. Sci. Gottingensis Recentioris* 6:3–50

Gibson JJ. 1950. *The Perception of the Visual World*. Boston, MA: Houghton-Mifflin

Gombrich EH. 1961. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. Princeton, NJ: Princeton Univ. Press

Gregory RL. 1968. Perceptual illusions and brain models. *Proc. R. Soc. B* 171:279–96

Hilbert D, Cohn-Vossen S. 1952. *Geometry and the Imagination*, transl. P Nemenyi. New York: Chelsea

Hildebrand Av. 1893. *Das Problem der Form in der Bildenden Kunst*. Strassburg, Ger.: Heitz & Mündel

Higashiyama A, Shimono K. 2012. Apparent depth of pictures reflected by a mirror: the plastic effect. *Atten. Percept. Psychophys.* 74(7):1522–32

Hoffman DD, Richards WA. 1984. Parts of recognition. *Cognition* 18:65–96

Hoffman DD, Singh M, Prakash C. 2015. The interface theory of visual perception. *Psychon. Bull. Rev.* 22:1480–506

Johnston A, Passmore PJ. 1994. Independent encoding of surface orientation and surface curvature. *Vis. Res.* 34:3005–12

Kennedy JM, Silver J. 1974. The surrogate functions of lines in visual perception: evidence from antipodal rock and cave artwork sources. *Perception* 3:313–22

Koenderink JJ. 2011. Gestalts and pictorial worlds. *Gestalt Theory* 33:289–324

Koenderink JJ. 2013 .... to see or not to see .... *Perception* 42:379–84

Koenderink JJ. 2014a. Q .... *Perception* 43:1015–17

Koenderink JJ. 2014b. The all seeing eye. *Perception* 43:1–6

Koenderink JJ. 2015a. Ontology of the mirror world. *Gestalt Theory* 37(2):119–40

Koenderink JJ. 2015b. PPP. *Perception* 44:473–76

Koenderink JJ. 2015c. Seeing with the ears. *Perception* 44:610–12

Koenderink JJ. 2015d. Visual art and visual perception. *Perception* 44:1–4

Koenderink JJ, van Doorn AJ. 1992. Surface shape and curvature scales. *Image Vis. Comput.* 10:557–65

Koenderink JJ, van Doorn AJ. 1995. Relief: pictorial and otherwise. *Image Vis. Comput.* 13:321–34

Koenderink JJ, van Doorn AJ. 2012. Gauge fields in pictorial space. *SIAM J. Imaging Sci.* 5:1213–33

Koenderink JJ, van Doorn AJ, Christou C, Lappin JS. 1996a. Perturbation study of shading in pictures. *Perception* 25:1009–26

Koenderink JJ, van Doorn AJ, Christou C, Lappin JS. 1996b. Shape constancy in pictorial relief. *Perception* 25:155–64

Koenderink JJ, van Doorn AJ, Kappers AML. 1992. Surface perception in pictures. *Percept. Psychophys.* 52:487–96

Koenderink JJ, van Doorn AJ, Kappers AML. 1994. On so-called paradoxical monocular stereoscopy. *Perception* 23:583–94

Koenderink JJ, van Doorn AJ, Kappers AML. 1995. Depth relief. *Perception* 24:115–26

Koenderink JJ, van Doorn AJ, Kappers AML. 1996c. Pictorial surface attitude and local depth comparisons. *Percept. Psychophys.* 58:163–73

Koenderink JJ, van Doorn AJ, Kappers AML. 2006. Pictorial relief. In *Seeing Spatial Form*, ed. MRM Jenkin, LR Harris, pp. 11–36. New York: Oxford Univ. Press

Koenderink JJ, van Doorn AJ, Kappers AML, Todd JT. 1997. The visual contour in depth. *Percept. Psychophys.* 59:828–38

Koenderink JJ, van Doorn AJ, Kappers AML, Todd JT. 2001. Ambiguity and the 'mental eye' in pictorial relief. *Perception* 30:431–48

Koenderink JJ, van Doorn AJ, Kappers AML, Todd JT. 2004. Pointing out of the picture. *Perception* 33:513–30

Koenderink JJ, van Doorn AJ, Pinna B, Pepperell R. 2016a. Facing the spectator. *i-Perception* 7:1–29

Koenderink JJ, van Doorn AJ, Pont S. 2012a. Shading, a view from the inside. *Seeing Perceiv*. 25:303–38

Koenderink JJ, van Doorn AJ, Wagemans J. 2011. Depth. *i-Perception* 2:541–64

Koenderink JJ, van Doorn AJ, Wagemans J. 2012b. Picasso in the mind's eye of the beholder: three-dimensional filling-in of ambiguous line drawings. *Cognition* 125:394–412

Koenderink JJ, van Doorn AJ, Wagemans J. 2013. SFS? Not likely! *i-Perception* 4:299–302

Koenderink JJ, van Doorn AJ, Wagemans J. 2014. Local shape of pictorial relief. *i-Perception* 5:188–204

Koenderink JJ, van Doorn AJ, Wagemans J. 2015a. Local solid shape. *i-Perception* 6:1–15

Koenderink JJ, van Doorn AJ, Wagemans J. 2015b. Part and whole in pictorial relief. *i-Perception* 6:1–21

Koenderink JJ, van Doorn AJ, Wagemans J, Pinna B. 2016b. Shading and the landmarks of relief. *Art Percept.* 4:295–326

Langer SK. 1953. *Feeling and Form: A Theory of Art*. New York: Charles Scribner's Sons

Lipschitz R. 1864. De explicatione per series trigonometricas instituenda functionum unius variabilis arbitrariarum, et praecipue earum, quae per variabilis spatium finitum valorum maximorum et minimorum numerum habent infinitum, disquisitio. *J. Reine Angew. Math.* 63:296–308

Maxwell JC. 1870. L. On hills and dales. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* 40:421–25

Meinong A. 1899. Über Gegenstände höherer Ordnung und deren Verhältnis zur inneren Wahrnehmung. *Z. Psychol. Physiol. Sinnesorgane* 21:182–272

Phillips F, Todd JT, Koenderink JJ, Kappers AML. 2003. Perceptual representation of visible surfaces. *Percept. Psychophys.* 65:747–62

Pizlo Z. 2001. Perception viewed as an inverse problem. *Vis. Res.* 41:3145–61

Pollack P. 1955. A note on monocular depth perception. *Am. J. Psychol.* 68:315–18

Reichel FD, Todd JT, Yilmaz E. 1995. Visual discrimination of local surface depth and orientation. *Percept. Psychophys.* 57:1233–40

Richards W, Bobick A. 1988. Playing Twenty Questions with nature. In *Computational Processes in Human Vision: An Interdisciplinary Perspective*, ed. ZW Pylyshyn, pp. 3–26. Norwood, NJ: Ablex Publ. Corp.

Rogers B, Graham M. 1982. Similarities between motion parallax and stereopsis in human depth perception. *Vis. Res.* 22:261–70

Rogers LR. 1969. *Sculpture: The Appreciation of the Arts/2*. London: Oxford Univ. Press

Rothe R. 1915. Zum problem des Talwegs. *Sitzungsberichte Berliner Math. Gesellschaft* 14:51–69

Sachs H. 1990. *Isotrope Geometrie des Raumes*. Braunschweig, Ger.: Vieweg

Saunders JA, Backus BT. 2006. Perception of surface slant from oriented textures. *J. Vis.* 6(9):3

Schlosberg H. 1941. Stereoscopic depth from single pictures. *Am. J. Psychol.* 54:601–5

Schrödinger E. 1958. *Mind and Matter* (Tarner Lectures, 1956). Cambridge, UK: Univ. Press

Schwartz AH. 1971. Stereoscopic perception with single pictures. *Opt. Spectra* 9:25–27

Spelke ES, Lee SA. 2012. Core systems of geometry in animal minds. *Philos. Trans. R. Soc. B* 367:2784–93

Spivak M. 1975. *A Comprehensive Introduction to Differential Geometry*. Boston, MA: Publish or Perish

Stevens KA. 1983. Surface tilt (the direction of slant): a neglected psychophysical variable. *Percept. Psychophys.* 33:241–50

Stevens KA. 1990. Constructing the perception of surfaces from multiple cues. *Mind Lang*. 5:253–66

Stevens KA, Brookes A. 1987. Probing depth in monocular images. *Biol. Cybernet.* 56:355–66

Stevens SS. 1957. On the psychophysical law. *Psychol. Rev.* 64:153–81

Strubecker K. 1942. Differentialgeometrie des isotropen Raumes III: Flächentheorie. *Math. Z.* 48:369–427

Teller DY. 1984. Linking propositions. *Vis. Res.* 24:1233–46

Todd JT, Koenderink JJ, van Doorn AJ, Kappers AML. 1996. Effect of changing viewing conditions on the perceived structure of smoothly curved surfaces. *J. Exp. Psychol. Hum. Percept. Perform.* 22:695–706

Todd JT, Reichel FD. 1989. Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychol. Rev.* 96:643–57

Todd JT, Thaler L, Dijkstra TMH. 2005. The effects of field of view on the perception of 3D slant from texture. *Vis. Res.* 45:1501–17

Todd JT, Thaler L, Dijkstra TMH, Koenderink JJ, Kappers AML. 2007. The effects of viewing angle, camera angle, and sign of surface curvature on the perception of three-dimensional shape from texture. *J. Vis.* 7(12):9

Topper DR. 1983. Art in the realist ontology of J.J. Gibson. *Synthese* 54:71–83

Tyler CW. 1975. Spatial organization of binocular disparity sensitivity. *Vis. Res.* 15:583–90

Vallortigara G, Zanforlin G, Pasti G. 1990. Geometric modules in animal's spatial representation: a test with chicks (*Gallus gallus domesticus*). *J. Comp. Psychol.* 104:248–54

van Doorn AJ, Koenderink JJ, Wagemans J. 2011. Light fields and shape from shading. *J. Vis.* 11(3):21

Vico G. 1725. *Scienza Nuova*. Naples, Italy: Muziana

Vishwanath D. 2014. Towards a new theory of stereopsis. *Psychol. Rev.* 121:151–78

Vishwanath D, Girshick AR, Banks MS. 2005. Why pictures look right when viewed from the wrong place. *Nat. Neurosci.* 8:1401–10

von Foerster H. 1992. Ethics and second order cybernetics. *Cybern. Hum. Knowing* 1:9–20

von Rohr M. 1904. *Linsensystem zur einaugigen Betrachten einer in der Brennebene befindlichen Photographie*. Kaiserliches Patentamt Patentschrift Nr.151312 Klasse 42h

von Uexküll J. 1909. *Umwelt und Innenwelt der Tiere*. Berlin: Verlag von Julius Springer

Wagemans J, van Doorn AJ, Koenderink JJ. 2010. The shading cue in context. *i-Perception* 1:159–78

Wijntjes MWA. 2017. Ways of viewing pictorial plasticity. *i-Perception* 8:1–10

Yuille A, Kersten D. 2006. Vision as Bayesian inference: analysis by synthesis? *TRENDS Cogn. Sci.* 10:301–8

Zeiss C. 1907. *Instrument zum beidäugigen Betrachten von Gemälden u.ggl.* Kaiserliches Patentamt Patentschrift Nr.194480 Klasse 42h Gruppe 34

Annual Review of
Vision Science

Volume 4, 2018

Contents

Errata

An online log of corrections to *Annual Review of Vision Science* articles may be found at http://www.annualreviews.org/errata/vision