Abstract. Depth is the feeling of remoteness, or separateness, that accompanies awareness in human modalities like vision and audition. In specific cases depths can be graded on an ordinal scale, or even measured quantitatively on an interval scale. In the case of pictorial vision this is complicated by the fact that human observers often appear to apply mental transformations that involve depths in distinct visual directions. This implies that a comparison of empirically determined depths between observers involves pictorial space as an integral entity, whereas comparing pictorial depths as such is meaningless. We describe the formal structure of pictorial space purely in the phenomenological domain, without taking recourse to the theories of optics which properly apply to physical space—a distinct ontological domain. We introduce a number of general ways to design and implement methods of geodesy in pictorial space, and discuss some basic problems associated with such measurements. We deal mainly with conceptual issues.

Keywords: depth, range, pictorial depth, depth scales, pictorial vision, pictorial relief, monocular stereopsis, depth cues, geometry of pictorial space, geometry of visual space, shape.

1 Introduction

“Depth”, as used in the experimental psychology of visual perception (Gibson 1950; Palmer 1999), is conventionally defined as the subjective correlate of “range”, where range is the distance of a fiducial object to the vantage point. (1) For simplicity we consider only the monocular observer. Then the vantage point is—for all practical purposes—the center of rotation of the eyeball (Graham 1965; von Helmholtz 1856). It is conventionally regarded as the task of vision to coordinate depth with range as well as possible. This is the mainstream ideal of veridical perception (Marr 1982). Perhaps unfortunately, this turns the bulk of depth-related vision research into a normative, rather than descriptive, enterprise.

There are conceptual problems with this view. For instance, as Wittgenstein (1921) observes in Tractatus (our translation):

5.632 The subject is not in the world, but is a boundary of the world.
5.633 Where in the world does one find a metaphysical subject? You say, it is much like the case of the eye and the visual field. But you don’t really see the eye. No feature of the visual field allows you to conclude that it is seen from an eye.

One arrives at the paradoxical situation that depth is supposed to be a correlate of range, whereas the eye and the object are not even together in visual awareness. The mainstream

(1) One often uses “distance” instead of range. We refrain from this in order to avoid certain inconsistencies. For instance, distances in three-dimensional Euclidean space are properties of nonoriented line segments, whereas range is a property of a direction which is obviously oriented. This is a categorical difference. The literature rarely makes the distinction.
view is that of an external observer. It fails to apply to the content of visual awareness, which is necessarily a first-person account.\(^2\)

A truly phenomenological account of depth—from the inside out, as it were—is required in order to arrive at the beginning of a psychological theory.

For the case of pictorial space this is the only rational starting point, because the notion of range is irrelevant to visual awareness. Only for the case of *vision-in-the-world* does one also require the relation to range. We have studied this extensively in our earlier work (Hecht et al 1999; Koenderink and van Doorn 1998a; Koenderink et al 2000; 2002a, 2002b; 2003; 2008). The theory of physical range is already in place. It is Euclidean geometry (for Euclid, see Burton 1945). What are needed in order to handle depth are forms of psychophysical bridge hypotheses and ways of operationalization of various aspects of depth.

There is a pervasive trend to consider pictorial vision as a limiting case of vision-in-the-world (Gibson 1950). This involves treating the picture as a window on a physical scene—so-called “Alberti’s window” (Alberti and Grayson 1972; Pirenne 1970). This would indeed put the eye in pictorial space (e.g., at the “perspective center”) and thus reduce all problems of pictorial perception to generic visual perception. Although thus highly attractive to the mainstream, it is a limiting, artificial case that can be approached only in laboratory settings. It is too restricted to be of much interest in the study of pictorial vision per se.\(^3\),\(^4\)

In this paper we attempt to construct an account of depth from the inside out, and we consider bridging hypotheses and aspects of various operational definitions of depth.

## 2 Depth

### 2.1 Depth as feeling

Phenomenologically, perhaps the most basic meaning of depth is as a general feeling of *separateness from the self*, or *remoteness*. One is indeed tempted to say “remoteness from the self”, but in immediate visual awareness the self does not necessarily figure at all. Immediate awareness is of the nature of “presentation”, something just happening, out of voluntary control, and prereflective.

In this sense remoteness, or separateness, occurs in all modalities, though perhaps in some more readily than in others. In humans it most prominently occurs in vision and audition, to a lesser extent in touch, and to an even lesser extent in gustatory and olfactory awareness. It is likely to be different in your cat or dog, and so forth. We can only guess with respect to sonar in bats, electropereception in sharks, and the like (Nagel 1974).

---

\(^2\) Even this is saying too much because visual awareness does not imply the notion of an ego.

\(^3\) We find repeatedly (Koenderink et al 2009) that human observers treat the “visual directions” as being lines perpendicular to the picture plane, thus all mutually parallel. Notice that this would place the eye at “minus infinity”—that is to say, outside pictorial space proper.

\(^4\) A good operational definition of the pictorial mode is to move your head or body. If you move somewhat in front of a painting, nothing much changes in pictorial space. In contradistinction, if you move your head when you presume to look into a window, you will notice peculiar, nonrigid deformations. This is a common experience when viewing baroque ceiling paintings. These are often of great extent and at considerable distance. If well painted, then they easily beat any “virtual reality” setup we have ever seen. The famous example (due to Pirenne 1970) is Pozzo’s ceiling painting in the church of *Sant’Ignazio* at Rome, but Europe (especially Germany, Austria, and Italy) is full of examples like this. Such deformations signify that your visual awareness is that of a window to which your eye is attached. On the other hand, the absence of such deformations when viewing a painting shows that your visual awareness is that of looking at a picture. Your eye is attached to the picture plane, but not to the pictorial space.
The feeling of separateness is perhaps induced through the feeling of *being acted upon*. Action and suffering\(^{(5)}\) are common to all life forms, including the most primitive ones (Schrödinger 1992).

### 2.2 Ordinal depth

Visual awareness can be more or less articulate. As one looks into a Ganzfeld the awareness is that of a luminous foggy atmosphere (Metzger 1953). It is indeed "out there" (remote), but the remoteness has no “value” (in a numerical sense). In perhaps more common cases visual awareness tends to be more articulate. In such cases the mere feeling of remoteness admits to degrees and can often be correlated with a scale.

If you articulate the visual field, for instance by presenting a statistically uniform arrangement of polka dots instead of a uniform field (Ganzfeld), most observers become aware of a surface (Koenderink et al 2009). The surface is apparently at a particular depth, because it is “thin” and “located”, but it is entirely undetermined what that depth is. The very notion of “(absolute) location in depth” is alien to the awareness. The surface just is—who knows where exactly?

From a formal, geometrical point of view one concludes that depth seems to have a single dimension, no obvious limits, and no origin (anchor) or unit (yardstick). A suitable model might be the geometrical line without additional properties.

That depth admits of something like a “location” becomes clear when you add various “cues”. By cue we mean any aspect of image structure that happens to influence the depth experience, not necessarily admitting of a physical or physiological “explanation” (Berkeley 1709). For instance, if we—in the aforementioned example—vary the individual sizes or colors of the polka dots, one tends to become aware of modulations of depth. The simplest example is the basic figure–ground distinction. The figure is experienced as “in front of” the ground. This may lead to arbitrarily long sequences of depths (Figure 1); thus it introduces an *ordinal depth scale*.

![Figure 1](image.png)

*Figure 1.* The picture at left is usually seen as a series of square tiles, stacked in depth. This depth order is roughly as shown in the volumetric view at center. The red arrow indicates the direction of view. In the alternative configuration at right one has a stack of rectangular tiles, with the right—instead of the left—one nearest to the viewer. This is another interpretation of the same picture at left. Both center and right interpretations of the picture at left are equally “valid” (both yield the projection shown at left), although the depth orders are mutually reversed.

From a formal, geometrical perspective the depth line becomes near-/far-polarized and admits of a serial order of points.

---

\(^{(5)}\) Being acted upon, perhaps through the encounter of a resistance to an action, may be denoted “suffering”, because it is not self-initiated.

\(^{(6)}\) In our experiment one out of eighty observers became aware of a volume filled with spheres, whereas all others reported to see a surface.
Notice that such a depth order is defined only *locally*. Globally, the depth order (if any) is only a partial order (Figure 2). Thus it is possible to speak of the depth order (which may or may not be defined or may be ambiguous) of *points in the visual field*, though not necessarily of a depth order of extended objects.

![Figure 2](image)

**Figure 2.** The red ring is neither fully in front nor fully behind the blue beam. Locally, both cases occur.

### 2.3 Depth scales

In many cases one is aware of a depth order that appears to be more strongly developed than just ordinal. “Absolute depth” is a meaningless notion in any case, but it is often possible to mutually compare depth differences, and perhaps there is even a notion of depth difference magnitude.

From a formal, geometrical point of view this implies that one has an *affine scale*. This involves a numerical measure, say a real number, characterizing the depth. Since the depth domain has no obvious boundary, this number may be taken to range from minus to plus infinity. For any given point two observers will most likely come up with different numerical depth values. But this does not mean anything in the absence of a natural origin and unit measure. One might capture this by saying that the depth is only relevant modulo(7) an arbitrary offset and unit (see Figure 3). Then the depth values assigned by the two observers no longer count as distinct.

![Figure 3](image)

**Figure 3.** The notion of “affine line”. The red point has coordinate 0.5 in the upper, but 1.666… in the lower scales. These scales are fully equivalent. They have different origins and units. Every affine property works equally well in either scale. For instance, the midpoint of the segment defined by two arbitrary points can be calculated (by taking the average) on either scale. Although the numbers (the coordinates) will differ, the point (as located on the corresponding scale) will be the same.

At the close of the 19th century the German sculptor Adolf Hildebrand (1901) noticed that human observers have difficulty discriminating work “in the round” from work “in relief”.

(7) Modulo is a standard technical term used in formal discussions. It means “disregarding certain variations regarded as irrelevant”. For instance, a geometrical configuration modulo position, size, and spatial attitude is “shape”.
In his ground-breaking book *The Problem of Form* he argued that the actual depth relief is largely irrelevant to human vision, though—of course—not to touch.\(^8\)

From a formal, geometrical point of view this implies that an augmented transformation that not only adds an arbitrary shift but also scales depth by an arbitrary (but positive) amount leaves the description invariant. This means that the depth domain is the “affine line”—that is, the real numbers modulo an arbitrary scaling and offset.

In the 1990s we (empirically, and only by accident!) discovered another basic ambiguity (Koenderink et al 2001). Human observers largely agree in their depth judgments modulo a transformation of an even further augmented type than with only scaling and offset—which is, with additional changes of the apparent frontal plane. In many cases their numerical depth judgments fail to correlate in a straight regression, whereas a multiple regression including *the picture plane coordinates*, perhaps surprisingly, reveals that the correlations are typically very high! This implies that the depth scales differ systematically for various points in the picture plane.

Each point of the picture plane carries the full depth dimension. We denote these depth dimensions “depth threads”. Thus *pictorial space is a sheaf of depth threads*. The structure of the picture plane serves to parameterize (label) them. Each thread carries its own ordinal depth scale. The finding mentioned above implies that these depth scales are different (a straight regression of depths reveals no correlation), though somehow coordinated (a multiple regression including the label “which thread” yields a high correlation).

Consider a plane in pictorial space that meets every thread just once. Any generic plane will do. Such a plane implies a depth value at every point of the picture plane. The results of depth measurements for different observers typically differ by exactly such an “additive plane”. It is as if observers may take different planes to be the (or their) “frontal plane”.\(^9\) This is an unexpected, and from a mainstream perspective surprising, finding.

A geometrical way to put this is to say that observers impose a “gauge” (or “gauge field”) on their pictorial space that can be represented as a pair of mutually parallel planes (see Figure 4). Such a gauge imposes an origin and unit point on all threads. It serves to mutually coordinate the depth scales on the various threads. Each observer (or the same observer at different occasions) applies an idiosyncratic gauge. As a consequence one cannot merely consider depth as such (that is to say, per visual direction), but one is forced to consider pictorial space as an integral entity. This has very important consequences for the analysis of psychophysical results. Perhaps unfortunately, the literature has been slow to take this up.

A naive analysis will use a default gauge that assigns the picture plane as base plane, conventionally denoted frontal plane. Unfortunately, observers often do not comply. Moreover, different observers make different choices. Failure to pick up on this is an important cause of trouble with literature results.

A gauge has four degrees of freedom. It can be thought of as composed of a shift in depth (one parameter), a scaling of depth (another parameter), and an additive plane (an additional pair of parameters). For zero shift, unit scaling, and a fronto-parallel additive plane one obtains simply the identity.\(^10\) If you apply two of these transformations in succession (a concatenation of gauges), you obtain just another gauge. Formally, one says that the gauge

---

\(^8\) The reason is that Hildebrand (1901) defines “pure vision” (or “artistic vision”) as vision based on a single viewpoint. In contrast, binocular vision and motion parallax are more like touch in the sense of providing multiple (simultaneous or successive) viewpoints.

\(^9\) This representation of the depth domain is formally identical to the “bas-relief ambiguity”, sometimes called “relief scaling and additive plane”, of computer vision (Belhumeur et al 1999).

\(^10\) An identity is a transformation that leaves everything the same. Such transformations are conventionally designated “identities” in formal discussions.
Figure 4. A “gauge” is a pair of parallel planes, a “base plane” B (in red), and a “unit plane” U (in blue). The gauge induces a scale in every depth thread (such as aa′ and bb′), the intersection with the base plane being the “origin” and the intersection with the unit plane the point at unit distance from the origin. Notice that the gauge serves to mutually relate the depth scales of all threads. The gauge induces different origins on the threads and synchronizes their units. In this picture the depth dimension extends vertically; the top and bottom planes of the box suggest a default gauge, which is parallel to the picture plane. The points on the thread have unit spacing. They serve to define depth numerically (as illustrated on aa′ and bb′), though for each thread individually. Each thread has its own scale as defined by the gauge.

Transformations form a group, which may be called the group of “depth similarities”. If the scaling is constrained to be one, you obtain a subgroup of “depth motions”. The group of depth motions is commutative; it does not matter in what order you apply motions, the result is always the same.

The gauge imposed by observers was already intuited by Hildebrand (1901), who suggested that one should conceive of depth as of the relation of an object with respect to two mutually parallel glass panes. This is just an implementation of a gauge. The insights of Hildebrand are surprising in view of the fact that modern science (early 21st century) has still to catch up with him!

2.4 The abacus model

The group of depth motions by definition leaves the structure of pictorial space invariant, by exactly the same philosophy that underlies the notion of “congruency” in Euclidean space: if two configurations can be superimposed through a motion, then they are considered to be geometrically identical. Notice that changes of gauge (the depth similarities) conserve a depth thread (visual direction) as a whole; they merely shift depth values along the threads.

This gives rise to an intuitive model of pictorial space that we will refer to as the “abacus model”. Remember that an abacus is a frame with a set of parallel wires on which you may shift beads. In the model we usually consider only a single bead per wire. The wires signify the depth threads, or visual directions, whereas the beads indicate depth values. Shifting the beads mimics the process in microgeny that establishes pictorial depth. Thus this part of microgeny (Brown 2002) appears much like a Glasperlenspiel (Hesse 1943).

---

(11) It is not hard to work out the nature of multiplication in the group, and one easily constructs the inverse of a gauge. The inverse always exists, and is unique.

(12) The full group of similarities is not commutative (or Abelian), though.

(13) “Microgenesis” is the term in phenomenological analysis that designates the process that generates visual awareness. It remains itself unconscious and is known only through its result.
Notice that each location of the visual field\textsuperscript{(14)} carries its own thread (depth direction) and that all threads are mutually isolated from each other. Thus the wires of the abacus are mutually unconnected and never merge. If we push all beads to their resting position (thereby “forgetting” all depth values), we “project” pictorial space on the picture plane. The artist who—for compositional purposes—views his or her painting as a quilt of colored patches applies such a projection. This feat takes practice, and most naive observers (nonartists) never achieve it.\textsuperscript{(15)}

The abacus model is evidently some kind of three-dimensional space. The essential difference between the abacus model and, for example, the familiar Euclidean three-dimensional space is that the visual directions are mutually isolated. Thus you cannot rotate about a line in the picture plane, as this would “mix up visual directions”. Such a rotation would reveal the back of the head of a portrait painted \textit{en face}, thus such a constraint certainly makes good sense. But remember that you \textit{can} apparently (mentally) “rotate” the apparent frontal plane! It merely takes a change of gauge. Fortunately, this is a non-Euclidean transformation that does not mix up visual directions, as will be shown below. In the abacus model the picture plane is “given”, whereas the microgenetic process underlying visual awareness shifts depth values along the visual directions as beads along the strings of an abacus. In this Glasperlenspiel such shifts are supposedly induced by the various cues identified by the microgeny (Figure 5).

\textbf{Figure 5.} An impression of the “abacus model” of pictorial space. The yellow “base space” is the picture plane; the red spheres indicate some fiducial locations in the picture plane. The gray vertical rods are the “wires” of the model; they stand for the depth threads (visual directions) associated with the points. The blue spheres indicate the “beads” of the model; they can be shifted to indicate depth values. Each bead slides only over its own wire. Here microgeny has placed the beads in the position of an additive plane. This “cross-section” indicates a pictorial relief (which is flat in this case!).

The final presentation, and the cues on which it is based, develop together, in close synchrony. Thus one cannot say that the presentation is a result of some algorithm applied to

\textsuperscript{(14)} We let the picture plane double for the visual field for the sake of this paper.

\textsuperscript{(15)} Formally, the abacus model is a “fiber bundle” over the Euclidean plane (“base plane”, here the picture plane), the “fibers” (the threads are technically known as “fibers”) being the affine line (the visual directions) and the “bundle projection” the projection mentioned here (Husemöller 1994).
some set of pre-established cues, which is perhaps closest to the current mainstream notion. Rather:

there is no presentation without cues, but neither are there cues without the presentation.

This is an essential point that tends to be ignored in mainstream accounts.

The geometrical entities that make up pictorial space are defined as the invariants of the transformations (Klein 1893) in the abacus model discussed above. That is to say, they have an existence that is independent of the actual description in terms of depth per se. They are defined in terms of what remains untouched after you subject that description to arbitrary transformations of the group of motions or similarities. Empirically, we have encountered frequent examples where observers of the same picture did not agree at all in terms of their depths, but very well in terms of the implied geometrical entities (Koenderink et al 2001). This should be crucial in any account of pictorial perception. A simulated example is shown in Figures 6 and 7.

2.5 Geometry of pictorial space

Much additional geometrical structure can easily be grafted on the abacus model.

Regard two distinct points of pictorial space. We define their “proper distance” as the familiar Euclidean distance of their projections on the picture plane. This makes sense, because it is obviously invariant with respect to the group of depth similarities.

But what if this distance vanishes? Then the points might still be distinct! Such points we call mutually parallel. In such a case of “parallel points” we define their “special distance” as the depth difference of these points. (We use quotes to indicate that these are not Euclidean terms.) Notice that this makes sense only if the proper distance vanishes, for otherwise a depth motion could change this special distance. Finally, we define “the” distance as either the proper or the special distance, according to which one applies. This distance is evidently an invariant of any pair of points, whether parallel or not.
Figure 7. At left the depth values corresponding to the cases illustrated in the previous figure (Figure 6) are shown in a scatter plot. Notice that the additive plane has destroyed the correlation, as the coefficient of variation is rather low. At right a multiple correlation that includes the position in the picture plane. This serves to “correct for” the different gauge. It is seen that the two depth distributions are after all identical. Gauges often suffice to “explain” the differences in depths for different observers. In such cases we say that the observers are aware of the same depth configuration and differ in only their “mental perspective” on it. Failure to notice this would lead to the (erroneous!) conclusion that the observers have fully different geometrical structures in visual awareness.

The shift parameter does not affect the proper distance, and acts the same on all wires of the abacus. It may thus be interpreted as a depth “translation”. However, the additive planes do act differently on different wires. We will show that they are naturally interpreted as (non-Euclidean) rotations. Finally, Hildebrand’s relief parameter can be interpreted as a kind of scaling, because it changes all special distances by the same factor. A “change of gauge” implies both a motion (translation and rotation) and a scaling.

The slant difference of two lines with the same projection in the pictorial plane may be defined as the special distance of any parallel points on them, divided by the proper distance of these points to the intersection of the lines (Figure 8). In this definition a rotation adds the same amount to the slopes of all lines that project on the same direction in the picture plane. This justifies both the definition of “rotation” and of “slant” (that is, non-Euclidean angle, Figure 9; Sachs 1990; Strubecker 1941, 1942, 1943, 1945; Yaglom 1968, 1979).

The slant of a plane is defined as the maximum slant of any line contained in it, whereas the tilt of a plane is defined as the direction in the picture plane of that maximally slanted line.

A pure rotation leaves a certain line in the picture plane invariant (which may be called the axis of rotation) and changes the slopes of all lines perpendicular (in the picture plane) to the axis by the amount of the rotation.

Notice that slants span the range of minus to plus infinity, thus the angle measure is not periodic as in Euclidean space. As mentioned earlier, this forces pictorial objects always to show their same side. You can never see their back sides, no matter what rotation you apply.

\[ R^2 = 0.684 \quad R^2 = 1.000 \]

---

\(^{(16)}\) We say “kind of scaling” because one may also scale the proper distances by changing just the size of the picture. Thus pictorial space allows of two distinct types of scaling. This is different from the Euclidean case because Euclidean angles—being periodic—cannot be scaled.
Figure 8. The angle, or slant, in the abacus model. Consider a point A (here in the base plane that does not matter) and two lines u and v through A. We consider “the angle between u and v”, but consider a wire that meets the lines u and v in P and Q, respectively. Let the common projection of P and Q be B. Both P and Q have proper distance AB (in red) to the point A, whereas their special distance is PQ (in blue). We define the angle subtended by the lines (in blue) as PQ/AB. As explained in the text, this angle is not periodic (because continued turning will not bring an object to its initial position), but falls within the range of minus to plus infinity.

Figure 9. A protractor for the abacus model. The angles run from minus to plus infinity; the angle measure is not periodic. This figure is the analog of a Euclidean wheel (shown at right). “Turning the wheel” in the abacus model means translating the vertical scale rigidly along itself; this corresponds to a rotation. Different from the Euclidean wheel, which repeats its attitude after a full turn, the non-Euclidean wheel cannot turn around. This geometry is in many respects similar to Euclidean geometry. However, the differences are crucial.

Patches of the visual field\(^{(17)}\) often occur as pictorial relief in visual awareness—that is to say, as coherent surfaces in pictorial space.\(^{(18)}\)

Surfaces often occur in awareness as the boundaries of coherent (volumetric) regions. For instance, a treetop is such a globular volume. Such surfaces occur as “smooth” on some level of detail, but typically (like the treetop) will show up additional detail that appears

\(^{(17)}\)Notice that we do not differentiate between the visual field and the picture plane in this paper. We would be more careful in an extensive discussion.

\(^{(18)}\)For the geometer such surfaces are technically known as “cross-sections” of the fiber bundle.
as “roughness” in the next finer resolution, and perhaps disperse in a flock of subobjects (leaves and twigs) on an even finer level of resolution. Such boundaries are two dimensional and therefore match the dimension of the visual field. They usually act as proxies for the volumetric regions themselves, since many pictorial objects are opaque.\(^{(19)}\)

Technically, a pictorial relief in terms of the abacus model is a surface that meets every wire only once. If such a surface is not planar, then one may attempt to assign it a “shape”. The shape should be invariant with respect to proper movements—that is, transformations with unit (that is, no) scaling but arbitrary translation and rotation.

Reliefs are typically patches of limited extent. If the reliefs are boundaries of volumetric objects, then the relief boundary can be of two categorically different types. One is a “contour”, which is the projection of a “rim”. A rim is a locus of points where the tangent planes have infinite slope. Such tangent planes are degenerate in the abacus model; you cannot even produce them on a physical abacus. For smoothly bounded volumetric regions the rim is a smooth, closed, twisted space curve. (See Figure 10, where the rim is planar.)

![Figure 10. The simplest example of a contour. The plane at the bottom is the picture plane; depth runs upwards. For clarity, most of the wires of the abacus model have been omitted. Only half of the orange surface of a globular volume is visible. The projection of the visible region in the picture plane is bounded by the yellow curve. This is the contour. It is the projection of the red curve—the boundary of the visible part of the surface, called the rim. The right figure is simply a rotated version of the left figure: the rim is different; the contour is identical. In these figures the “invisible part” has been partly omitted. It is not “optically specified”, but is present in awareness, especially near the rim: it becomes gradually more uncertain the farther it is removed from the rim. “Present in awareness” means that the observer (prereflexively) entertains certain expectations. This can be demonstrated—in the case of vision-in-the-world—by rotating the object in physical space, thus revealing previously hidden parts. When expectations fail to meet, the observer will show surprise, thus revealing the presence of the expectation.]

The contour is a curve that may contain self-intersections and cusps. The contour as projection of the rim is somewhat of an “ideal” object, in that not all parts of it are necessarily visible (Koenderink and van Doorn 1982).

The other type of boundary is an “occlusion boundary”. It happens where another object occludes the fiducial one. In a sense the occlusion boundary does not belong to the fiducial object, and in visual awareness the fiducial object is represented as passing behind the occluder. These simple cases are slightly complicated through the possible existence of self-occlusions. Occlusions proper and self-occlusions imply that the visible contour is limited

\(^{(19)}\) Such regions may be called “amodal” in the sense that they are somehow “posed” by the microgeny, but are not represented at all in the optical structure. One is always aware of the backsides of objects (they are “real” in some sense). This can even be shown experimentally (van Lier and Wagemans 1999).
to only part of the projection of the rim. Visibility toggles at “T-junctions” and cusp points (Figures 11 and 12).

Figure 11. The simplest example of a T-junction. The plane at bottom is the picture plane, depth runs upwards. For clarity, the wires of the abacus model have been omitted. Only half of each of the yellow-green, checkered surfaces is visible. The projections of the visible region in the picture plane have been colored brownish and greenish. In this case one of the surfaces is partly occluded by the other. Thus the brown region is bounded by contour alone, whereas the green region is bounded partly by its contour, partly by the occluding contour, which is the contour of the other surface. The single depth direction that has been drawn is tangent to both surfaces. Its projection is a so-called “T-junction”.

Figure 12. The “ending” contour. The plane at bottom is the picture plane; depth runs upwards. For clarity, the wires of the abacus model have been omitted. Only part of the greenish, checkered surface is visible because it “folds over itself”. The rim is a piecewise smooth curve. In this case its projection in the picture plane has a sharp “cusp”. Only one branch of the cusp is visible, the other branch being occluded by the surface itself. Thus one sees a contour that ends at the red point; the projection of the point on the surface that marks the end of the pleat.

In the interior of a pictorial relief the depth is defined at any point; thus one has a continuous depth field. In visual awareness this is the essence of “surfaceness”—smooth surfaces and smooth depth fields being one and the same. This increases the number of local, depth-related quantities that may—in principle—be addressed via psychophysical methods. The most important ones are local depth proper, the spatial attitude of the local tangent plane, and the local shape.
In the simplest case one studies the shape in the immediate neighborhood of a point—this is conventionally known as its curvature. The simplest way to do this is to apply a transformation that rotates the local tangent plane so as to be fronto-parallel. You may also apply a translation, both in depth and in the picture plane, to move the point to the (arbitrarily assigned) origin. Near the origin you study the shape of the curves of equal depth (Koenderink and van Doorn 1998b; Figure 13). Generically, these will be either concentric ellipses or concentric hyperbolae, corresponding to elliptic (convex or concave) and hyperbolic (saddle shaped) surface curvature. These quadrics are known as “indicatrices of Dupin” (Koenderink 1990) and geometrically specify the local curvature. The indicatrix of Dupin has the shape of the “wound” inflicted to a region if you cut its surface with the blade held parallel to the tangent plane at the fiducial point. For instance, cutting chips off a sphere produces all circular wounds, illustrating the fact that the spherical surface is curved identically in all directions (Koenderink 1990).

Figure 13. (a) At left a relief in the abacus model. At center the curves of equal depth in the picture plane. Suppose the red point is currently of interest: at right we show the level curves of depth in the immediate neighborhood of the point. Notice that the point is “located on a slope”. (b) This continues the situation illustrated in (a). We applied a depth motion such that the tangent plane of the fiducial point becomes fronto-parallel. At left the rotated relief in the abacus model. At center the curves of equal depth in the picture plane. Remember that the red point is currently of interest. At right we show the level curves of depth in the immediate neighborhood of the point. Notice that the point is no longer located on a slope. It has turned into an extremum of depth. The level curves have become more or less concentric ellipses (the “indicatrix of Dupin” at the fiducial point), whose sizes and shapes represent the local curvature at this point. Generically one obtains either ellipses (like here), representing convexities or concavities, or hyperbolae (saddle shapes).

3 Geodesy in pictorial space

To be more precise: for a given depth value the numerical curvature is inversely proportional to the diameter of the indicatrix in the direction of curvature. The directions of principal curvature are the axes of the indicatrix. From this one easily obtains the “shape index” (Koenderink and van Doorn 1992).
How to practice geodesy in pictorial space? This is obviously an important topic in visual psychophysics. It is also a very elusive one because pictorial space is a mental entity. You cannot introduce the familiar tools of the geometer—such as yardsticks, carpenter’s squares, calipers, or what have you—into this space. In fact, you cannot introduce anything physical into pictorial space, you can introduce only pictorial objects—that are mental entities—into pictorial space.

Fortunately, it turns out to be quite easy to put pictorial objects into pictorial space. You simply superimpose the picture of the object over the fiducial picture. If the picture of the object is much smaller than the fiducial picture, then the object will be sucked into the pictorial space evoked by the latter until it hits the nearest obstacle, to which it attaches itself. This is a matter of common observation. For instance, if you draw a black blotch on the poster of some person, it will attach to the pictorial person. Examples include beauty spots on cheeks, black teeth, moustaches, and so forth on the pictorial faces of politicians and movie stars. If you are a city dweller, you will be only too familiar with this. The trick is simply to put the blotch at roughly the suitable place in the picture plane. A little practice will soon give you a feel for the basics. The actual structure of the mark is hardly important, as a little exercise on a portrait of your most-hated politician will soon teach you.

Once you are able to put whatever you fancy into pictorial space, the gate is opened widely to a variety of geodesic methods. We will discuss a few in this section, though rather cursorily, since the main thrust of this paper is to discuss the essential differences between such methods, especially of methods that all purport a single geometrical property—say depth. This is—or at least it should be—a major issue in psychology because if one depth is ontologically distinct from the next one, then the difference should not be the cause of mere controversy but should give rise to attempts at integration and improved understanding. Perhaps unfortunately, this is still a distant target.

3.1 Varieties of geodesy

What can be measured at all, and what might be viable methods of praxis of geodesy in pictorial space? In this subsection we take a quick look at both issues.

3.1.1 What can be measured?. One way to arrive at a toolbox is to consider geometrical entities of various dimension, as well as their interrelations. In pictorial space one distinguishes points, curves (lines, orbits), surfaces (planes, reliefs), and volumetric regions (solid shapes). Points are characterized by depth; the simplest geometrically relevant entity is the special distance. When points are close their depth difference may be of interest. Lines have slants and curves have changing slants—that is, curvature. Slants may be changed by rotations, whereas curvatures are invariant. The incidence of a point and a line or curve is nongeneric, thus remarkable. Planes and reliefs have spatial attitudes that may be changed by rotations; reliefs have invariant “curvature landscapes”. The incidence of reliefs with points or curves is nongeneric, thus remarkable. Curves generically intersect reliefs in points, and two reliefs generically intersect in a curve. Volumes of limited size may or may not contain certain points, and so forth. This yields a rich set of possible relations to explore. Any of these is a potential target for research. We have been able to sample only a few instances thus far.

An important distinction is between submanifolds, which are smooth curves or surfaces, and configurations of discrete elements, such as sets of mutually independent points or planes. In the case of submanifolds one meets with important constraints. For instance, if you sample local surface attitude (eg, slant and tilt) along some closed curve on the surface, then you should arrive at the same attitude after traversing the full curve. Such “surface integrability conditions” play an important role in the formal theories of cues such as shape from shading. Moreover, the continuity allows such basic operations as smoothing and interpolation, which are very important in the statistical analysis of empirical data. In the
case of discrete configurations you may also meet constraints, but only if you measure more parameters than are needed to specify the configuration. The latter condition is common enough, though. For instance, if you measure the depth relation (closer or more remote?) for all point pairs in a point configuration, you sample many more responses than there are depths. Geometrical constraints are very valuable in permitting a check on the consistency of empirical data; thus they play an important, often crucial, role in the analysis.

How to measure? There are numerous possibilities, although these possibilities have remained largely unexplored.

Methods may be distinguished in a variety of ways. One is whether the method addresses local or multilocal properties. Notice that “local” may involve more structured entities than mere location; think of the spatial attitude of tangent planes or curvature, for instance. Another, important, distinction is whether the “probes” applied to the picture merely indicate spatial presence or mimic some geometrical structure, allowing a nontrivial notion of “fit”. In this section we offer some examples of the various possibilities, though by no means exhaustively so.

One of the simplest methods is based on the pairwise comparison of entities. For instance, given two points, one may ask which one is closer (van Doorn et al 2011). This technique may be almost arbitrarily articulated. For instance, one may ask whether a given point lies in front or behind the plane implied by a triple of other points, and so forth. The simple comparison judgment is subject to enormous variation. It includes (important) judgments of incidence. Any of these methods depends on only the identification of points. They can be implemented by putting simple marks (say colored dots) on the picture plane. These dots simply serve to indicate a location—there is no notion of “fit”.

Another generic method is that of fitting a “gauge figure”. The idea is simply that in order to measure something (anything!), you compare it with a standard. For instance, you measure the length of a body by putting a yardstick next to it, you measure the weight by comparing it with a standard weight using scales, you grade grains by means of sieves, and so forth. In all such cases the observer merely has to judge the fit of a pictorial entity with respect to some “gauge object”.

For such methods you need to overlay a picture of the gauge object over the fiducial picture. In many cases you will need to grant the observer control over the picture of the gauge figure. A simple example involves an elliptical gauge figure to measure the spatial attitude of a planar patch of relief (or the tangent plane at a point of a pictorial relief; Koenderink et al 1995). Here the task is to judge whether the gauge figure appears as a “circle painted on the pictorial surface in pictorial space” (Figure 14). This samples surface attitude, usually parameterized by the slant and tilt angles. Notice that this parameterization is fully irrelevant in the actual perceptual judgment. The observer need not even be familiar with these parameters. A fit is a fit when it looks like one, and that is all there is to it. When a fit is established, the result of the measurement is taken to be the current gauge in some parameterization. This is simply an operational definition. This is how the ancients measured the weights of things, although they were sorely lacking a reasonable theory of gravitation or mass. It is not essentially different in psychophysics.

This general technique of judging the fit of some gauge object is capable of almost boundless variation and can be used to measure a large variety of geometrical properties.

---

(21) The method of fit is familiar to physicists: It has been the major method of quantification in classical physics. It is often misunderstood in cognitive science, though. One speculates that the observer judges both the object to be measured and the gauge figure separately, and subsequently judges whether these two partial judgments agree. The idea ultimately derives from the (mistaken) notion that everything mental has to be cognitive. This again leads to attempts to calibrate the gauge figure by some independent method. From the perspective of the physicist this is nonsensical.
Figure 14. Portrait of an angel by Albrecht Dürer (completed 1506) with superimposed gauge figure. This particular gauge figure is made up of the picture (in orthographic projection) of a circle with an “axle” sticking out at right angles from the plane of the circle. The axle has the same length as the radius of the circle. The task is to make the picture of the circle, which is elliptical due to foreshortening, look like a circle “painted” on the surface, in this case the angel’s cheek. If the observer is satisfied, the spatial attitude of the gauge figure is taken as a sample of the tangent plane of the relief (the cheek) located at the center of the gauge figure.

Yet another method involves multilocal properties (notice that the gauge figure methods are essentially local). One introduces objects at two or more distinct locations and lets the observer judge (or adjust) a multilocal fit. A simple example involves pointing from one point to another, obviously an important operation in practical geodesy. In this case one overlays a picture of a pointer on one location and a picture of a target on the other (Figure 15). The observer then adjusts the picture of the pointer so as to “point from one point to the other” in pictorial space (Wagemans et al 2011). That this should be possible is evident from the fact that one easily follows visual directions of pictorial persons. For instance, in paintings by Peter Paul Rubens the pattern of “who looks at whom” in sizable groups of persons tends to be very well established. It was an important part of the mature baroque style (Andersen 1969).

Another example of a multilocal method deploys “depth cues” such as relative size or atmospheric perspective. One simply superimposes a circular blob (say) over each location and lets the observer adjust either relative size or relative color such that the blobs (they tend to look like “spheres”) look “the same” in pictorial space (Figure 16; van Doorn et al 2011). These methods are also capable of considerable variation and development.
Figure 15. A scene from Sergio Leone’s *Once Upon a Time in the West* (a well-known movie from 1968), with superimposed target and pointer. (In an actual application these superimposed objects will be smaller.) The observer controls the spatial attitude of the pointer and is supposed to make it “look as if it points at the target in pictorial space”. When the observer is satisfied, the attitude of the pointer is taken as the direction in pictorial space from the location of the pointer to that of the target.

### 4 Some empirical results

#### 4.1 Relation between various differential depth orders

In the case of depth reliefs one has a variety of possible geometrical entities at one’s disposal that might be sampled empirically. For instance, consider

- sampling the depth order for pairs of points on the surface;
- sampling the spatial attitude of the local tangent planes;
- sampling the local curvatures.

In these cases one measures different properties of the same surface; thus the resulting data are expected to stand in certain relations to each other. Of course, the empirical relations are likely to be predictable from the formal, geometrical relations.

If the depth of all surface points is known (a “depth field”) up to a common offset, then the surface attitude can be calculated through spatial differentiation. Conversely, if the spatial attitude is known throughout (a “surface attitude field”), then the depth field can be calculated by integration. There is a possible hitch, though. An arbitrary field of attitudes need not be integrable (need not be a “gradient field”). This is up to empirical verification. If integrability fails, then no consistent pictorial relief can be said to exist.

Similar considerations apply to the relations between depths, attitudes, and curvatures.

It is a priori unclear in what way—if any—surfaces are represented in visual awareness. This issue can be addressed empirically. For instance, if the depth field can be predicted from the attitude field, but not vice versa, then this is an indication that not depths per se, but rather attitudes are represented.

---

(22) Formally, this involves the various orders of differential description. Depth is the zeroth order, attitude the first order, curvature the second order, and so forth.

(23) Notice that this would count as a valuable empirical result.
Figure 16. A capriccio (imaginary landscape) by Francesco Guardi (1712–93) with two yellow blobs superimposed on it (copy by Anne-Sophie Bonno; see http://www.atelier-bonno.fr/galerie-copies-arts-graphiques.html). Notice that they tend to look like spheres in pictorial space. The task is to adjust their relative size so as to make them “look the same size in pictorial space”. If the observer is satisfied, we take the logarithm of the size ratio as a measure of the depth difference between the locations identified by the blobs.

In one study we found that attitudes are represented more precisely than depths and that the attitude field is a gradient field (Koenderink et al 1992, 1996). We are still uncertain whether curvatures might take priority over attitudes, which remains certainly a possibility.

If one samples depth priority for nearby locations, one finds results that can be predicted from attitude samples. When the points are located far from each other, one might expect their depth priority to be predictable via integration of the attitude field. We find empirically that such is not the case, though. It works in case the points are located on a single slope, but it fails if there runs a depth ridge, or rut, inbetween them (Koenderink and van Doorn 1995). We conclude that observers have no immediate access to the integrated attitude field at all. Notice that this would be the global depth field. This is the case even though the depth field was obtained from the observer’s own responses, which is perhaps surprising. Such findings rather immediately address the geometrical structure of visual awareness.

Relations such as these are very intricate, and science is still far removed from a comprehensive understanding of them. Even so, such an understanding must be considered to be of paramount importance to our understanding of pictorial vision.

4.2 Comparison of operationally defined depths

The simplest depth scale is ordinal. One distributes a number of points over the pictorial scene and collects pairwise depth ranking judgments. For a reasonable number of points (roughly fewer than 50) one may let the observer judge all (orderless) pairs within an hour or so. For coherent pictures (we tried a classical landscape) one finds that the final ranking is coherent, and stable over sessions (van Doorn et al 2011). The resulting rank order typically (over a number of observers) has about 40 distinctly different levels. Thus observers notice quite an articulate depth structure in such a landscape picture.

(24) The time limit is important because observers tend to apply different transformations (rotations and scalings) from day to day.

(25) This is not trivial since 50 points implies that 1,225 \((50 \times (50 – 1)/2)\) judgments are “explained” by a mere 49 \((50 – 1)\) degrees of freedom.
In a method of relative size comparison the observer has a more complicated task. One again tests on pairs of locations, each trial involving the adjustment of relative size of probes. We find that a session may comprise about 20 locations, implying 190 adjustments (van Doorn et al. 2011). Again, we find a consistent structure, both within and over sessions.

Finally, in a pointing task the observer has the even more complicated task, involving two degrees of freedom, of pointing a pointer to a target. In this case one can use only fewer than 10 locations (we used 5, involving 20 trials). In all cases one points once either way, because it turns out to be the case that the observer “points by arcs” rather than straight lines. The result is again consistent within (27) and over sessions (Wagemans et al. 2011).

The latter two methods yield numerical scales, the former one only an ordinal scale. We find that the rank correlation between all methods is very high, and the same is found for the Pearson correlation coefficient between the numerical scales; this is the case for all observers. We do find high correlations between observers—apparently they all use more or less the same cues—but very significant differences between the depth ranges obtained with the latter two methods (van Doorn et al. 2011). On first blush this appears surprising, given the fact that all observers appear to resolve a similar number of distinct depth levels.

Consider the relative size method. What we did was to take the natural logarithm of the size ratio adjusted by the observer as a measure of the depth difference between the two fiducial locations. The rationale for this choice is simple enough. When you scale both probes by the same factor, the depth difference of the corresponding locations is not supposed to change. Hence the depth difference may depend upon only the ratio of the sizes. Moreover, the depth differences should combine additively, whereas size ratios combine multiplicatively. Hence one should consider the logarithm of the ratio. The base of the logarithm is arbitrary, and thus one obtains the depth up to some unknown factor (the same for all observers), and an arbitrary offset. The latter is irrelevant—we arbitrarily set the average of the depths obtained over the session to zero.

Notice that the depth scale obtained in this way is unrelated to the dimensions as measured in the picture plane. This is fine, because the proper depth motions of pictorial space conserve the visual directions; thus there is no principled way to compare the magnitude of distance in the picture plane and depth. What one may do is find the slant of the line that connects the points in pictorial space. It is the depth difference divided by the distance of the points in the picture plane. This slant is subject to the relief ambiguity described by Adolf Hildebrand. All slants are ambiguous because they may be arbitrarily scaled with a common factor. The total depth range is subject to the same ambiguity. Since the common factor must be assumed to be essentially idiosyncratic, this is not necessarily at odds with the fact that all observers resolve similar numbers of depth levels.

The relative size cue has nothing to do with any “ratio of ranges”—after all, the eye is not even in pictorial space. The observer somehow uses the size cue “smaller is farther away”, even in the absence of a clear “distance from the eye”. It seems a priori likely that this will be highly dependent upon the particular style of the picture. For instance, in Persian miniatures—where the size–range relation is apparently ignored, and the major depth cues are height in the picture plane and overlap—things might well be expected to work out quite differently. There lies a vast field of enquiry open here.

Next consider the pointing method. It differs essentially from the relative size method in the respect that the measure of distance of the points in the picture plane does matter. In

---

(26) This is not trivial because one attempts to account for the 190 \((20 \times [20 – 1]/2)\) adjustments in terms of a mere 19 \((20 – 1)\) degrees of freedom.

(27) This is not trivial. One accounts for twenty \((5 \times [5 – 1])\) settings in terms of four \((5 – 1)\) degrees of freedom.
this case one measures the depth difference between two locations in terms of their mutual distance in the picture plane. If one abstracts from the fact that the two-way pointings yield slightly different results, then the depth difference is simply the distance in the picture plane times the slant adjusted by the observer. This slant is defined as the tangent of the Euclidean slant angle used in the computer graphics generation of the picture of the pointer. Notice that this is intuitively reasonable in view of the definition of the non-Euclidean slant in pictorial space. Whereas the Euclidean slant angle is in the range of minus to plus ninety degrees (positive meaning away from the observer), its non-Euclidean analog ranges from minus to plus infinity. The tangent is just a number, whereas one measures distances in the picture plane in pixels, visual angle, or centimeters—they have units. Then the depth difference will also be expressed in pixels, visual angle, or centimeters.

Does this mean anything?

Yes and no. In one possible view the result is again subject to Adolf Hildebrand’s relief scalings. Thus one draws similar conclusions as in the case of the relative size probe. The results of both methods may be normalized by dividing by the total range or depth standard deviation before comparison. When one does so, the results are almost indistinguishable.

However, another view is possible in which one puts more value on the habitus of the pointer as a function of its slant.

The picture of the pointer is constructed via standard computer graphics. It is the orthographic projection of a three-dimensional arrow. The shaft and head of the arrow have been dimensioned in such a way that small variations in slant are readily apparent in any spatial attitude of the pointer. When looking at the arrow in various attitudes, one can roughly estimate its slant angle in degrees. In doing this one of course has to assign a depth scale. It seems likely that one does this on the basis of the implicit assumption (which is correct) that the arrow is a rotational symmetric body in Euclidean three-dimensional space, seen in orthographic projection. When the image of the pointer is superimposed over the fiducial image, the arrow is perceived as a pictorial object in the pictorial space of the fiducial picture (a classical landscape). One may still look at the pointer and readily estimate its slant angle in degrees.

Does this imply that observers setting different slant angles experience different depth differences? It seems to us that one may take this as an operational definition that this is the case. Accepting this implies that the experienced depth range is independent of the depth resolution. One might say, then, that the pictorial depth calibration is a form of mental paint.

5 Depth and range

In this paper we have developed the topic of geodesy in pictorial space without any reference to the structure of physical space. We believe this to be the proper procedure, because a picture (in terms of its structural complexity) in no way depends upon any physical space. Here Maurice Denis’s (1890) manifesto (my translation), is very appropriate:

Remember that a painting—before being a battle horse, a nude woman, or some anecdote—is essentially a planar surface covered with colors in a certain arrangement.

This is trivially obvious in the case of Mondrian’s later paintings, but to many people it is rather less obvious in the case of “realistic paintings” or even photographs. Visual artists understand Denis’s statement very well, though, especially painters. The painter’s craft is to

\[28\] It can be shown that this problem is easily resolved by using the average slope of the two empirically determined slants.

\[29\] The tilts are simply discarded in this calculation as they involve only the picture plane per se.
apply pigments on a planar carrier in such a way as to trick their potential customers to build certain constrained hallucinations in their visual awareness. The planar arrangement is in no way superseded by these, though.

The “composition” has an important effect. That is explicit to the artist although it remains largely implicit to the naive observer. One might say that the observer is being brainwashed by the artist. We mean this quite literally. This is—in principle—not different from what happens in musical composition.

In the context of vision-in-the-world the relation between the space in awareness, and the physical space that contains the observer is of importance, of course. It then makes sense to look for possible correlations between properties of depth and those of range.

For the restricted setting of the static, monocular observer—only eye movements, no motion in the environment—the case is not that different from that of pictorial vision, the main difference being that “ground truth” is overwhelmingly present.

For this restricted case one may deduce the possible structure of visual space from first principles (Koenderink and van Doorn 2008). One predicts that depth is correlated with the logarithm of range and that the structure of visual space must admit a certain group of similarities that is essentially the same as we deduced above on phenomenological grounds. The exploration of the empirical structure of this visual space is far from completed yet. This poses a rich field of endeavor in its own right.

6 Conclusions

Pictorial vision is a niche area of vision research that remains relatively unexplored, either theoretically or empirically. This is the case despite the fact that this apparently limited setting occurs very frequently in our current society. Just think of the evacuation plans the airlines are legally forced to bring to your attention at the beginning of their flights. A considerable part of communication and teaching still depends heavily on material presented as still pictures. Moreover, much of our historical documents come as either texts or still pictures. There is no way technology might change this, for in the medieval authors simply had neither stereoscopes nor video. Thus a thorough understanding of pictorial vision, its dependence on the pictorial material, and the differences between members of the general population (perhaps graded with respect to age and gender) should be of considerable interest.

Unfortunately, the understanding of pictorial vision is typically regarded as a rather immediate derivative of the—immensely more important—understanding of vision-in-the-world. This is a consequence of the understanding of pictures as being essentially windows on some physical space. Such a view introduces the observer into the picture, thus converting the viewing of a picture as a limited instance of vision-in-the-world. For instance, the experiential quality depth is interpreted as the physical range. That is to say, one understands “veridical vision” as implying an intimate correlation between depth and range. In our view this is nonsensical, because of the ontological chasm between these concepts. Moreover, the normative emphasis is anthropocentric and clashes with a biological (thus evolutionary) account. However that may be, the fact remains that pictorial vision is a worthwhile field of scientific endeavor in its own right.

We know this may sound odd, but artists understand it very well. As one of many examples remember the famous dictions by Edgar Degas:

A picture is something which requires as much knavery, trickery, and deceit as the perpetration of a crime. Make it untrue and add an accent of truth . . . Even working from nature you have to compose . . . Drawing is not what you see but what you must make others see.
We have shown how a formal theory of depth can be developed through pure phenomenological deduction. We have also indicated how one may design and implement methods of geodesy that can be applied in the empirical investigation of structures in pictorial space. Such methods have hardly been explored in any detail thus far. There is ample room for further development.

Perhaps not unexpectedly, it is fair to say that a great many open problems remain, especially of a conceptual nature. Pictorial vision is quite different from vision-in-the-world because the conventional perception–action cycle does not apply to it. Thus the definition of vision as “optically guided behavior”, which is quite acceptable to much of the mainstream, fails to apply to it. It has to be changed into “optically constrained awareness”, which lifts it on a different ontological level. The study of pictorial vision has to rely on phenomenological methods, rather than physiological or “hard-core” psychological ones.\(^{31}\)

**Acknowledgements.** This work was supported by the Methusalem program by the Flemish government (METH/08/02), awarded to Johan Wagemans. We would like to acknowledge administrative support by Stephanie Poot and useful comments on a previous version by two anonymous reviewers.

**References**

Alberti L B, Grayson C, 1972 *On Painting* (New York, NY: Penguin Classics) ◄

Andersen L, 1969 *Baroque and Rococo Art* (New York, NY: Harry N Abrams Inc) ◄

Bellhumeur P N, Kriegman D J, Yuille A L, 1999 “The bas-relief ambiguity” *International Journal of Computer Vision* **35** 33–44 doi:10.1023/A:1008154927611 ◄

Berkeley G, 1709 *An Essay Towards a New Theory of Vision* (Dlin, UK: Aaron Rhames, at the Back of Dice Coffee-House, for Jeremy Pepyat, Bookseller in Skinner-Row) ◄

Brown J W, 2002 *Self-embodying Mind: Process, Brain Dynamics and the Conscious Present* (Barrytown, NY: Barrytown/Station Hill Press) ◄

Burton H E, 1945 “The optics of Euclid” *Journal of the Optical Society of America* **35** 357–357 doi:10.1364/JOSA.35.000357 ◄

Denis M, 1890 “Définition du néo-traditionnisme” *Art et Critique* **65** 556–558 ◄

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

Graham C H, 1965 *Vision and Visual Perception* (Oxford: John Wiley & Sons) ◄

Hecht H, Koenderink J J, van Doorn A J, 1999 “Compression of visual space in natural scenes and in their photographic counterparts” *Perception & Psychophysics* **61** 1269–1286 doi:10.3758/BF03206179 ◄

Hesse H, 1943 *Das Glasperlenspiel* (Zürich: Verlag Fretz und Wasmuth) ◄

Hildebrand A, 1901 *Das Problem der Form in der bildenden Kunst* (Strassburg: Heitz & Mündel) ◄

Husemöller D, 1994 *Fibre Bundles* (New York: Springer Verlag) ◄

Klein F, 1893 “Vergleichende Betrachtungen über neuere geometrische Forschungen” *Mathematische Annalen* **43** 63–100 doi:10.1007/BF01446615 ◄

Koenderink J J, 1990 *Solid Shape* (Cambridge, MA: MIT Press) ◄

Koenderink J J, van Doorn A J, 1982 “The shape of smooth objects and the way contours end” *Perception* **11** 129–137 doi:10.1068/p110129 ◄

Koenderink J J, van Doorn A J, 1992 “Surface shape and curvature scales” *Image and Vision Computing* **10** 557–564 doi:10.1016/0262-8856(92)90076-F ◄

Koenderink J J, van Doorn A J, 1995 “Relief: pictorial and otherwise” *Image and Vision Computing* **13** 321–334 doi:10.1016/0262-8856(95)99719-1 ◄

Koenderink J J, van Doorn A J, 1998a “Exocentric pointing” in *Vision and Action* Eds L R Harris, M Jenkins (Cambridge: Cambridge University Press) pp 295–313 ◄

Koenderink J J, van Doorn A J, 1998b “The structure of relief” in Advances in Imaging and Electron Physics (New York: Academic Press) pp 65-150 ◄

Koenderink J J, van Doorn A J, 2008 “The structure of visual spaces” *Journal of Mathematical Imaging and Vision* **31** 171–187 doi:10.1007/s10851-008-0076-3 ◄

Koenderink J J, van Doorn A J, Kappers A M L, 1992 “Surface perception in pictures” *Perception & Psychophysics* **52** 487–496 doi:10.3758/BF03206710 ◄

\(^{31}\)Such hard-core psychological methods are really to be understood as “noninvasive physiology”, rather than as psychology proper.
Jan Koenderink (1943) studied physics, mathematics, and astronomy at Utrecht University, where he graduated in 1972. From the late 1970s he held a chair “The Physics of Man” at Utrecht University until his retirement in 2008. He is presently Research Fellow at Delft University of Technology and guest professor at the University of Leuven. He is a member of the Dutch Royal Society of Arts and Sciences and received a honorific doctorate in medicine from Leuven University. Current interests include the mathematics and psychophysics of space and form in vision, including applications in art and design.

Andrea van Doorn (1948) studied physics, mathematics, and chemistry at Utrecht University, where she did her masters in 1971. She did her PhD (at Utrecht) in 1984. She is presently at Delft University of Technology in the department of Industrial Design. Current research interests are various topics in vision, communication by gestures, and soundscapes.

Johan Wagemans (1963) has a BA in psychology and philosophy and an MA and a PhD in experimental psychology, all from the University of Leuven, where he is currently a full professor. Current research interests are mainly in so-called midlevel vision (perceptual grouping, figure–ground organization, depth, and shape perception) but stretch out to low-level vision (contrast detection and discrimination) and high-level vision (object recognition and categorization), including applications in autism, arts, and sports (see http://www.gestaltrevision.be).