

Pictures: Crafting and Beholding

Jan Koenderink and Andrea van Doorn

University of Leuven (KU Leuven), Belgium
E-mail: koenderinkjan@gmail.com

Abstract. *The psychogenesis of visual awareness is an autonomous process in the sense that you do not “do” it. However, you have some control due to your acting in the world. We share this process with many animals. Pictorial awareness appears to be truly human. Here situational awareness splits into an “everyday vision” and a “pictorial” mode. Here we focus mainly on spatial aspects of pictorial art. You have no control whatever over the picture’s structure. The pictorial awareness is pure imagery, constrained by the (physical) structure of the picture. Crafting pictures and beholding pictures are distinct, but closely related, acts. We present an account from experimental and formal phenomenology. It results in a generic model that accounts for the bulk of formal (rare) and informal (common) observations. © 2023 Society for Imaging Science and Technology. [DOI: 10.2352/J.Percept.Imaging.2023.6.000401]*

1. INTRODUCTION

This paper sketches our time path in arriving at a formal understanding of “pictorial space”. We are writing in retrospect. So we are backed up with extensive empirical evidence gathered over decades. Hence we are able to outline a much more coherent perspective than was possible during the course of our quest.

We “discovered” relevant facts in arbitrary order and dealt with them conceptually as best as we could at the time. Now we are to trace interrelations and “post-dictions” that we unfortunately passed up as pre-dictions.

So we present a neat formal account that should provide a useful framework for further investigations.

1.1 Daily Vision and Pictorial Vision

“Everyday vision” is a process that uses optical structure — both “received” and “sought for” — to efficaciously act in the world. It may be regarded as a kind of *user-interface* [20, 27, 32]. For the most part human everyday vision is not essentially different from the vision of other animals. It is for a large part dynamic and autonomous. It involves the whole body¹ and a huge store of experience gathered over evolutionary timespans [33, 34].

“Pictorial vision” is different in that binocularity, parallaxes and other “physiological cues” serve only to reveal a picture as the *planar* (think of a painting on the wall) object

Received Oct. 24, 2022; accepted for publication May 3, 2023; published online July 26, 2023. Associate Editor: Michael A. Wester.

2575-8144/2023/6/000401/11/\$00.00

¹ It may be remembered that the eyes are mounted roughly on top of the platform and are mechanically connected to the world by the feet.

it is. A picture is not a window in that you don’t look *through* it, but *at* it, at least if you see it as a physical object [10]. However, many humans — in this we are different from animals — are able to look *into* a (2D-)picture and experience a virtual, or imaginary 3D space. This is not “inverse optics”,² but something science is — by definition — unable to handle.

Pictorial space is an enigma for the sciences. Indeed, it is frequently dismissed as self-contradictory since there is no relevant disparity. Most dictionaries define “stereopsis” as *binocular* stereopsis, whereas vision science sometimes refers to “paradoxical monocular stereopsis” [7]. There is hardly a reason to doubt the experience of pictorial (thus monocular) depth [3, 4, 22, 40]. No visual artist doubts it, whereas many vision scientists do.

This was the reason for our interest in pictorial vision since the early 1990’s. It involved a shift from physiology and psychophysics — in its proper sense — to experimental phenomenology [1]. Most of our efforts were spent on the development of novel methods to obtain precise, structured, quantitative data on pictorial qualities like shape and depth. This led to the discovery of lawful “holder’s shares” [15] that allow the essentially idiosyncratic optical awareness of different people to be formally accounted for.

Only *modulo* a group “mental movements” (see A.5.1) can such awarenesses be quantitatively compared. It led to a novel type of theoretical (or formal) phenomenology with considerable predictive power. Although not “science” in the strict sense, it opens up an academic endeavor with both formal and empirical underpinnings.

2. OPTICKS

We use the ancient spelling to distinguish our topic from *optics*. It is indeed fundamentally different from what you find in the “Optics” textbooks. The first text on opticks is by Euclid [13].

Euclid’s opticks is either construed as a defective theory of physical optics, or as a defective theory of linear perspective. It is neither,³ it is a geometrical information theory involving “visual rays”.

The theory has obvious applications to terrestrial animal vision, indeed so obvious that Euclid is rarely mentioned.

Bishop Berkeley [6] put his finger on the crux:

² The optics is not invertible because it is a projection that discards one dimension (range) [6].

³ It has nothing to do with the propagation of electromagnetic disturbances, nor with projections on picture planes [21].

... distance, ... cannot be seen. For distance being a line directed end-wise to the eye, it projects only one point in the fundus of the eye, which point remains invariably the same, whether the distance be longer or shorter.

Opticks involves a projection, which is why it is not invertible. Thomas Reid [37] most incisively researched the consequences in his phantasy of the Idomenians.

2.1 The Frontal Observer

Euclid and Reid considered a formal momentary point observer with a perfectly isotropic and homogeneous structure. However, humans are not like that. Our system (like most organic systems) is neither homogeneous, nor isotropic [42].

We are bipedal animals with an optical system that samples the half-space in front of us. So we'd better adopt a coordinate system that recognizes that. Here we use Cartesian coordinates, with the X -axis in the principal viewing direction. The canonical viewing direction is the (horizontal) forward direction.

Euclid's opticks is the same in all planes that contain the viewing direction. Thus — for the sake of succinctness (and intuition!) — we will discuss mainly the horizontal plane at eye height in this paper. It does not sacrifice generality.

Whereas Alberti's "linear perspective" [2, 8] appears to be a mere artistic convention, it actually fits the human condition quite well.⁴ So this is the preferred system: X -axis forwards, Y -axis horizontal to the right, Z -axis vertically upwards. Here we discuss and illustrate the XY -plane, the "horizontal plane". The sagittal (XZ) plane has the same geometry. The frontal (YZ) plane is special. It is the boundary of the "scene" and the space behind the observer (which is optically *elsewhere*).

The generalization from XY to XYZ is trivial and adds nothing new.

2.2 Specialization of the Euclidean Projection to the Human Case

We deal with what is essentially Alberti's perspective [2]. However, we prefer (because formally most simple) a perhaps unfamiliar representation. Overall, we prefer synthetic geometrical methods, but the conventional algebraic method has its uses.

The optical projection is implemented by a hyperbolic involution in projective space (see Appendices A.3.1 and A.3.2, [9], Figure 1). For a forward looking observer, we define the plane of the "viewport" as $x = 1$, thus we use the "viewing distance" as the natural unit of length. It is the distance (see appendix A.1) to the nearest relevant points of the scene and may vary greatly. The size of the viewport is nearly equal to the viewing distance for a "normal view", smaller for a "narrow angle" view and larger for a "wide-angle" view (see A.3.4).

⁴ The reason is generic biology, rather than something special due to our singular position relative to fellow animals [42, 43].

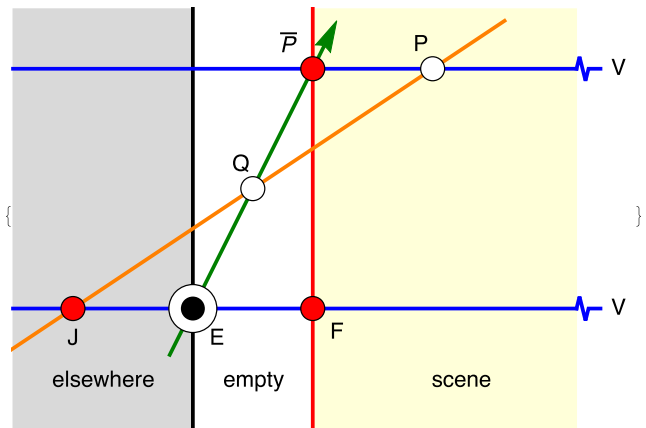


Figure 1. The hyperbolic involution that defines "perspective". The red objects (point J, and red line through the frontal point F) are invariant. Vertical lines and the lines through J are conserved as lines, but generically not point-wise. The points P and Q which are mutually swapped by the involution lie on the orange line through J. The origin is the "eye" (point E). Rays through the origin (green arrow) map on horizontal lines (blue), they meet at the principal vanishing point V at infinity. The involution swaps the eye and the principal vanishing point, and likewise P and Q. The constructions of P given Q and of Q given P work in exactly the same way.

On the X -axis (principal visual ray) we have that $u = 1/x$. Thus the *elsewhere* region $x < 0$ is mapped onto itself and so is the scene $x > 0$. Within the scene the map swaps the *empty* region $0 < x < 1$ and the *visual scene proper* $1 \leq x < +\infty$ ("proper" because it singles out the part of the scene in front of the observer). See Fig. 1.

The central projection on any frontal plane⁵ is nothing but Alberti's linear perspective (see A.3.2). Thus the projection contains Alberti's perspective, but retains the distance (x) order and maps it on a "nearness" ($u \in (0, 1)$) order. We refer to $1 - u$ as the "depth".

For a canonical view (size equals the viewing distance⁶) one has a field of view of approximately 53° . We refer to the image of the frustum as the "viewbox". For the normal view it is a cubical region. The visual rays in the viewbox are parallel to the principal view direction, the "eye" being swapped with the principal vanishing point (see A.3.1 and Figs. 1 and 2).⁷

Of course, the viewbox is a severely squashed version of the scene (Figures 3 and 4). However, it has at least *some* thickness, whereas the Albertian perspective is completely flat.

We interpret the projection in a special way. Its domain is *physical space* (coordinates $\{x, y\}$), whereas its range is a *mental space* (coordinates $\{u, v\}$), referred to as "visual space". Thus $\{x, y\}$ and $\{u, v\}$ are incommensurable

⁵ Useful choices are the frontal plane through the eye, the frontal point, or the principal vanishing point. One choice may be preferred over others according to the context of the discussion.

⁶ Examples are the "normal focal length" in photography and the traditional sizes of letter paper.

⁷ Formally, the viewbox is what is known as the NDC (Normalized Device Coordinates, or depth buffer) in Computer Graphics.

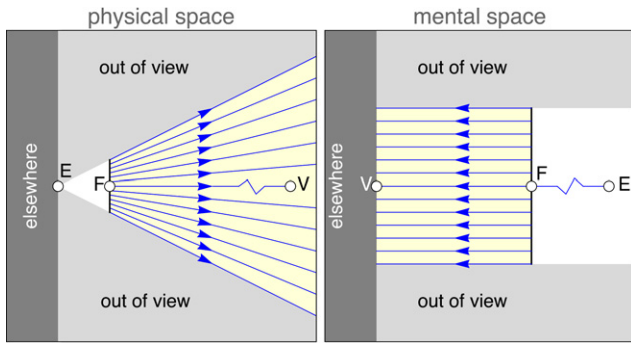


Figure 2. The basic optical projection. Here E denotes the eye, F the frontal point of the viewport, V the principal vanishing point. The “elsewhere” region is the space behind the back, the “out of view” region (which is connected!) is what is cut off by the viewport. Although not optically specified, such regions are in situational awareness. The white region is the empty space between the eye and the viewport. Visual rays are plotted in blue, note their direction! The zigzags signify an infinite interval. The yellow area indicates the frustum (left) and the viewbox (right). It may take some thought to get familiar with the topology.

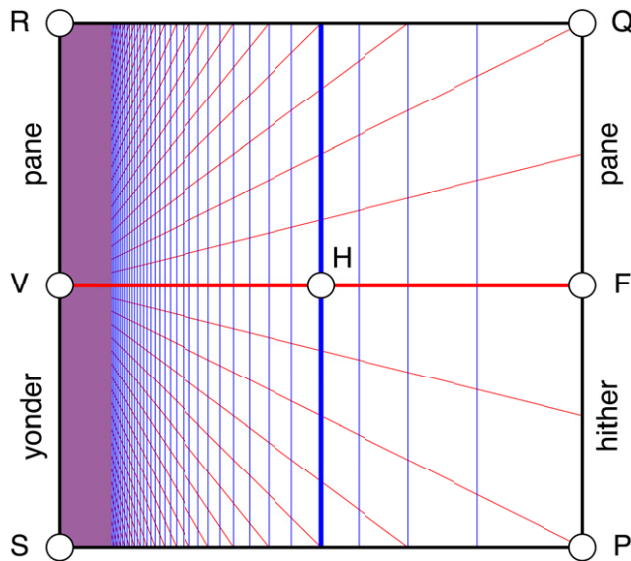


Figure 3. The viewbox is spanned by the boundary P, Q of the viewport and the boundary R, S of the field of view. The principal viewing direction (red line) runs from the frontal point F to the principal vanishing point V. The centre is the hyperfocal point H, the hyperfocal plane is $u = \frac{1}{2}$. The red lines are loci of constant y , the blue lines represent equidistance planes. Transversal scaling is u , longitudinal scaling u^2 , so things get smaller and flatter with depth. Note the shears in oblique directions.

because they are *ontologically distinct* [25]. This differentiates our interpretation from scenography, which treats the domain and range on ontologically identical terms [16]. The formal description is identical, but the meaning is categorically different.

2.3 A Further Specialization to Pictorial Vision

A picture is a plane covered with pigments in some simultaneous order [10]. Thus *there is no scene*. A picture does not come with “ground truth”, that would appeal to a God’s Eye view.

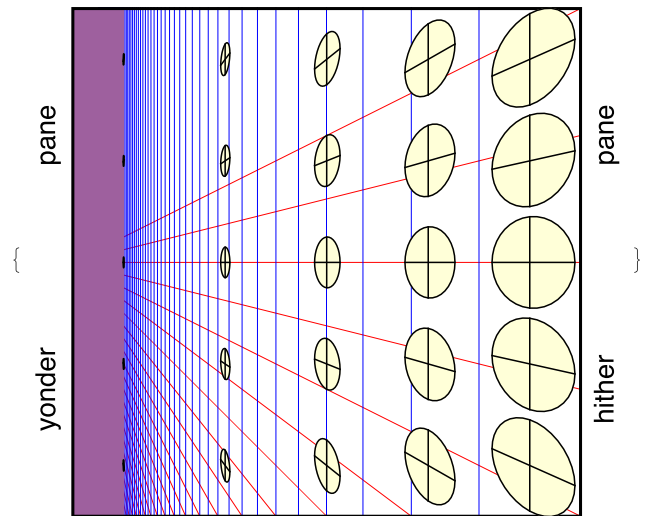


Figure 4. The metric can conveniently be judged by mapping circles of equal size. We indicate the images of diameters in the X- and Y-directions. From a formal standpoint one constructs the Riemann metric tensor. This also allows one to find the curvature (the viewbox is “flat”) and geodesics (straight lines with projective parameterization).

Since there is no scene, the viewbox cannot be the image of some (non-existing) frustum. It is a mental figment, some form of imagery. We refer to the space-frame of this imagery — in case it exists in awareness — as a “picturebox”.

In order to formalize this, we define the depth dimension in the picturebox as *isotropic* (see A.5), that is to say the full real line “rolled up in a point”. This allows for an (imaginary) 3D-space packaged in a $(2 + 1)D$ picture plane. This non-Euclidean “thick plane” is a well known geometrical object with attractive⁸ properties [38, 39, 41, 44, 45].

The group M of (special) similarities conserves the (physical) picture plane but affects the (imaginary) depth (see A.5.1 and Figure 5). Because these similarities conserve the picture plane point-wise, they are essentially arbitrary, although they conserve the geometry (the projective structure). Thus the picturebox should be taken *modulo* arbitrary similarities M . Pictorial observers may apply arbitrary M -transformations as “beholder’s share”. We refer to these as “mental movements” [28].

The group M largely coincides with the group of ambiguities for the familiar *Shape from X* algorithms. For instance, the so-called “bas-relief ambiguities” identified for shape from shading are elements of M (see A.6, [5, 23]).

3. EXPERIMENTAL PHENOMENOLOGY OF PICTORIAL SPACE

Here are the most essential findings distilled from a huge set of (quantitative and/or qualitative) empirical phenomenological observations:

⁸ Indeed, simpler than Euclidean, because the distance and angle metrics are both parabolic. In contradistinction, the Euclidean plane has an elliptic angle metric which leads to numerous pesky “exceptions”.

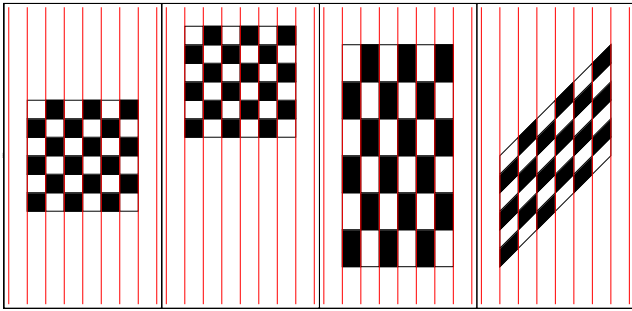


Figure 5. Examples of “mental movements”. The red lines are the wires of the abacus model (Figure 6). Starting with a fiducial configuration we have, from left to right: a translation, a scaling and a rotation. Note that only the depth is affected, the picture plane coordinates remain invariant.

- typically, observers yield (often *very*) different depth fields. However, applying suitable mental movements these differences often become rather minor. Depth correlations may change from non-significant to 0.99. Without such corrections the observations are essentially meaningless and will lead to faulty conclusions [28].
- observers relate spatial attitude to the local visual ray direction [24]. This explains a large part of the well known “deformations” of linear perspective renderings.
- observers assume a default “normal field of view” of 40°–60°. This explains another large part of the deformations often ascribed to linear perspective renderings [24, 29].
- observers experience “depth”, a projective image of *distance*, as an image of *range* (see A.1) [25].

These rules (so far) seem to have no exceptions and may (at least in this stage of the investigation) be taken as “Laws of Pictorial Perception”. We view them as an additional chapter to the well known Gestalt Laws [12, 26, 30, 35]. Of course, there are numerous useful facts of a different nature that might eventually make it to such a status.

The rules predict various huge(!) deviations from “veridical perception” — not even mentioned in traditional textbooks — in a quantitative way. They may also be used to arrange (or even deform) scenes and pose actors in such a way as to yield intended pictorial awareness in viewers of the resulting pictures [36].

4. A FORMAL MODEL OF THE PSYCHOGENESIS OF PICTORIAL SPACE

How does the creative imagination fill the picturebox so as to present one with a pictorial space? This is the problem of the psychogenesis of pictorial space. The process runs in pre-awareness, the results simply happen (like sneezing), one doesn’t “do” it in reflective thought. It is obviously “creative” (the opticks being non-invertible), but equally obviously constrained by the physical structure of the picture (pixels, paint strokes, or whatever). The relevant structures are conventionally referred to as (pictorial) “cues”.

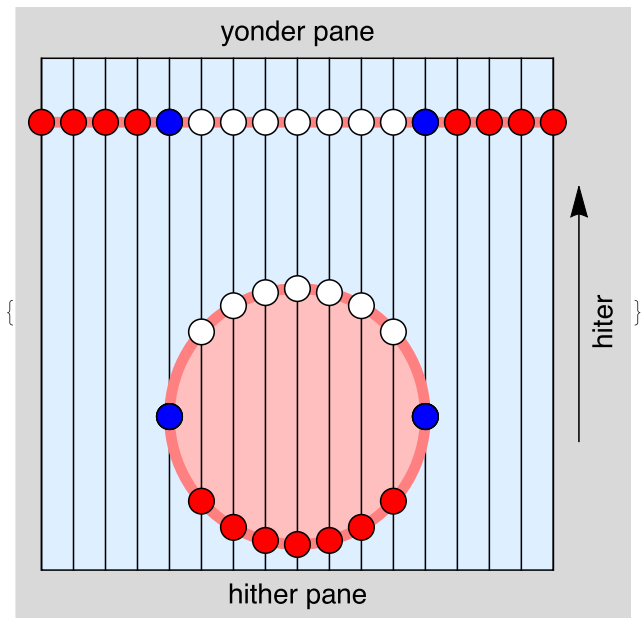


Figure 6. A simple example of a representation in the Abacus Model, here a convex object in front of a backdrop. Regular beads are shown in red, “virtual beads” in white and “split beads” in blue. The split beads are singular, here they are points on the occluding contour. The virtual beads are actually of different kinds. Some are on the backsides of objects, others are invisible due to occlusion with a nearer object.

Here we present a simple quasi-geometrical model of such psychogenesis. The model is generic in the sense that it does not cover the nature or use of specific cues, that would be a subsequent stage. Thus it leaves open the topic of whether pictorial vision is predominantly “constrained hallucination” or some form of (“regularized”) “inverse optics”. (Our choice would put the emphasis decisively on the former.) We refer to it as the “Abacus Model” (see A.7 and Fig. 6).

The picturebox is a rectangular volume⁹ with two opposite edges parallel to the viewing direction, thus taking account of the fact that all visual rays in the picturebox are mutually parallel. The box is capped by a hither pane and a yonder pane.

The *hither pane* is the mental image of the picture surface. For ease of reference we will mutually identify those, although it should be kept in mind that they are ontologically different. Then the visual rays may be labelled as “pixels”. Each pixel can be thought of as extended “in depth”. The depths are essentially arbitrary. For instance, when viewing a picture obliquely, one may notice the hither pane to break apart from the perceived physical picture surface [31].

The *yonder pane* is a “backdrop” at perceptual infinity, yet the “total depth” of the picturebox is experienced as finite. The yonder pane is perhaps best understood as the plane of vanishing points, or — what amounts to the same — the

⁹ Although we do not mention it repeatedly, remember that this immediately transfers to the 3D case. The 3D viewbox and picturebox are cuboids.

space of directions of visual rays in a (in this case imaginary) scene.

Typical pictorial objects are opaque surfaces of volumetric entities. One experiences the frontal sides of these, simultaneously being aware of their volumes and hidden backsides. Thus psychogenesis somehow assigned a depth-field along the rays belonging to an area of pixels. It is as if it shifted a bead along the ray until it was “stopped” by some opaque surface [19]. Thus one may conceive of psychogenesis as a “bead-shifting game”.¹⁰ This is indeed how it has been characterized by practicing visual artists [11, 17].

We refer to this model as the “Abacus Model”, because it is obviously reminiscent of the counting frames often used to teach children the arithmetic of addition. Note that there may be more than one bead on a thread, generically either one (a surface), two (an occluding contour), or three (a T-junction). There may also be “virtual beads”¹¹ representing the backsides of visual objects or otherwise occluded surfaces.

The group of mental movements M (see A.5.1) is essential in the experimental phenomenology of pictorial space. It by and large accounts for observer variations and the variations found for a single observer under repeated viewings.

5. SOME CONSEQUENCES OF THE FORMALISM

We mention just a few remarkable cases. Of course, the formalism has a much wider applicability.

5.1 Windows and Pictures

The view of a picture has again and again been compared with the experience of looking through a window. It describes the experience of pictorial space at least to some extent. However, it is very rare — taking very sophisticated measures — to *confuse* the experience of looking through a window with that of looking at (or even into) a picture. Such experiences (if any) are limited to professionally executed

¹⁰ See the introductory chapter of Hermann Hesse’s *Das Glasperlenspiel* [18].

¹¹ Remember that *all* beads are imaginary!

instances. Generically, when looking *into* a picture, you have the simultaneous situational awareness of looking *at* a picture surface. This is (very much!) part of the charm of pictorial perception. “Illusions” are the meal tickets of Madame Tussaud’s and Disneyland. They have little or nothing to do with enjoying (or even using) pictures [19].

The difference is due to the fact that you are not one of Euclid’s, Berkeley’s or Reid’s momentary point observers. You probably have both eyes open and move your body, head and eyes continuously. This immediately identifies a window as an *aperture* and a picture as an *opaque surface* in your situational awareness.

5.1.1 Oblique Viewing

Keyhole vision is an extreme case of looking through a window. The crux is that any vantage point yields a distinct view. For a picture, the case is different. The vantage point only deforms the optical structure a bit, but it does not add or remove anything.

Is it possible to “see the same scene” as you change your vantage point with respect to the window frame? Yes, it is, but you have to change the physical scene for that. A scene and some affine deformation of it may be set up so as to provide the same optical data. This was famously proven by de la Gournerie in the mid nineteenth century (see A.4).

The proof has often been used to explain certain “deformations” due to the oblique viewing of pictures [36]. This is nonsense for a variety of reasons, most importantly ontological.

When you view a picture obliquely, the abacus model predicts that your pictorial awareness will only change by a mental movement, whereas your situational awareness, including the awareness of the physical picture surface, will reflect the usual effect of a change of vantage point. Thus “constancy mechanisms” will apply to the latter, but not to the former. This has the following experiential effects (Figure 7)

- A pictorial face will “follow you”. The reason is simply that an en-face view will remain that, it cannot become a (say three-quarters) profile.
- The hither pane will break free of the apparent picture surface. Note that the two are in ontologically



Figure 7. Lord Kitchener points straight at the spectator. An oblique sideways view shows the head as thinner, but the pointing is not affected. A really weird viewpoint renders the picture illegible.



Figure 8. Bill Brand's (1961) *Nude on the Beach* extends far into the background. Its space is strongly dilated. In contrast, Mantegna's *Dead Christ* (ca. 1480) is just a shallow foreground with a featureless dark backdrop at indeterminate depth. Its space is strongly contracted.

distinct spaces, thus cannot formally be compared. Your awareness manages to deal with simultaneously present mutually distinct spatial frameworks.

- Oblique viewing leaves the apparent size of the picture frame invariant (“size constancy”), but the pictorial object will apparently shrink in one direction [31].

It is easy enough to check such predictions in front of a suitable picture with a few minutes of mindful observation.

5.2 Various Well Known Problems with Linear Perspective Renderings

Ever since its advent in the early Italian Renaissance, artists had to deal with complaints concerning “perfect” linear perspective renderings. By the end of the Renaissance artists had figured out how to beat these problems using various intentional deviations from the correct methods. However, there never was something like a formal canon for these.

We find that the bulk of the effects is captured by the simple assumption that psychogenesis treats the geometry of the viewbox and picturebox much as the space they move in. No doubt there will be various corrections to this model, that might account for the observations in more detail. However, we are convinced (on the basis of our own, rather extensive, observations over some decades) that this simplest model is surprisingly effective and generic.

More extensive and precise observations — unlikely to become available in the near future — will almost certainly turn up deviations, although these may well turn out to be of an idiosyncratic nature.

5.2.1 Apparent Rotations

Consider a row of mutually translated, but otherwise identical objects. Think of a row of persons lined up in perfect military order. If the field of view is say 90° or more, most observers spontaneously notice that the outermost persons appear to look to the side. They appear “rotated” [24].

The amount of apparent rotation is well predicted by the eccentricity, even up to eccentricities of a hundred degrees.

This is an immediate consequence of the Law of Locality. In order to make a photograph that “looks right”, all actors need to view the camera.

5.2.2 Apparent Sizes

Depth may reflect range or distance, depending upon current situational awareness. That is another reason why the Albertian perspective breaks down for wide-angle views.

In order to experience an arrangement of actors as being posed in perfect military order, they have to face the camera and be at equal range from the camera. But Albertian perspective translates this into weird pictorial size variations.

In order to “look right” one needs to use Guido Hauck's *Plattkarte* (“equirectangular projection”).¹² That this works very well can easily be demonstrated using a fish-eye¹³ instead of a regular photographic objective [24].

5.2.3 Apparent Frontal Planes

Frontal planes typically appear as curved with the convexity turned towards the eye. This may be due to the fact that one has to look back and forth in order to cover the horizon. The planar wall somehow disappears into depth at the horizontal vanishing points to the left and right. A concave circular wall with the eye at its centre actually *looks* frontoparallel. This is a consequence of the locality principle [24].

5.2.4 Apparent Depth Dilations and Contractions

If the viewing distance relative to the size of the main object is varied the perspective changes. The immediate impression is

¹² The *Plattkarte* is the simplest example. Many cylindrical projections work just as well and may have advantages in certain settings.

¹³ At least on the horizon, most fish-eye lenses are “angle true”. Full optical array cameras often yield perfect equirectangular maps.

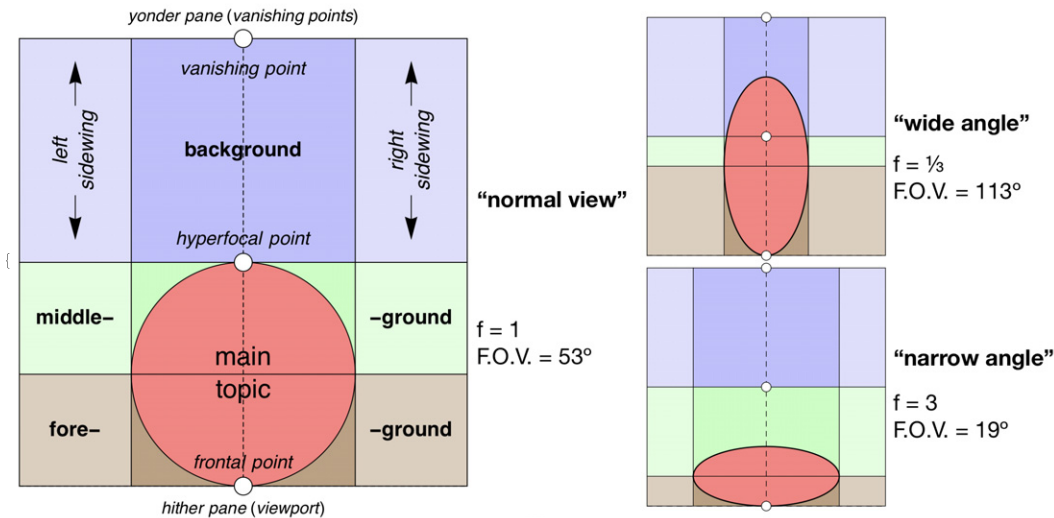


Figure 9. At left a typical layout: a roughly isotropic, convex “main topic” embedded in a scene. The frontal plane is placed such as to touch the main topic, so all of the setting is behind it. The main topic is given half the width of the viewport, leaving a quarter width for both wings. This layout lets us define the foreground, middle ground and background depth layers. In the “normal view” (viewing distance f equals width) the field of view is 53° . At right we show what happens for wide-angle (113°) and narrow angle views (19°). (Note that the viewport is the same in all cases.) This explains the effects seen in Figure 8.

that for a short viewing distance near objects become drawn out in depth (Fig. 8 left), whereas, for a long viewing distance they are flattened (Fig. 8 right).

The reason why wide-angle views appear to have an expanded space, whereas narrow-angle views appear to be contracted in depth is shown in Figure 9. (See appendix A.3.4.)

Perhaps strange enough, there really is something like a “canonical scene” (Fig. 9 left). It contains a major topic set in a scene. The major topic will fill at least half of the scope and is typically roughly isotropic (sphere-like or cube-like). This is crucial for the segmentation of depth.

If the main topic extends beyond the hyperfocal point (“halfway infinity”, see A.3.3) there is only a foreground which intrudes into the background. If the main topic does not reach the hyperfocal point there is room for a middle ground between the foreground and the background. The former case applies to wide-angle views, the latter to narrow-angle views (Figs. 8 and 9).

5.3 Effective Methods of Depiction

The various problems discussed above are easy to solve in practice. Apparent rotations are cured by counter rotation. Either all actors face the vantage point, or one uses orthographic projection. If an impression of “correct perspective” is desired, one uses the correct construction on the architecture, but treats the actors as seen frontally in orthographic projection.¹⁴

Apparent sizes are trivially corrected by adjusting pictorial sizes as indicated by eye measure.

¹⁴ A conventional example is Raphael’s fresco *Scuola di Atene* (1509–1511, Stanza della Segnatura, Apostolic Palace, Vatican City).

Apparent frontal planes are treated differently, depending upon the scope of the depicted scene and the planned field of view during viewing. The viewing situation may be “forced”, for instance by hanging a painting relative to an entrance and so forth.

Apparent depth dilations and contractions are easily controlled by juggling planned depth layers. For a typical rendering one may use up to three proper (having finite thickness) layers (foreground, middle ground and background) and a (flat) backdrop. Each layer is treated orthographically at a magnification that helps indicate its place in the depth order. So linear perspective is simplified and discretized in a well planned manner (Figure 10). This leads to renderings that “read well”, often better than a real scene would. A well crafted picture is like a “Reader’s Digest” version of the scene (as in Fig. 10).

In the scene the observer has to construct cues,¹⁵ whereas in the pictures the artist *communicates* them. This assumes a common cultural background of picture crafter and beholder.

6. CONCLUSION

We present a generic model of pictorial space based on a huge base of data from experimental phenomenology. The model is “generic” in the sense of leaving open all kinds of specializations, it is hardly more than a succinct formalization of general observations. So we do not consider it “speculative”, but we believe that it will be part of the framework of any future specialization.

¹⁵ Cues need to be constructed, rather than “found”, because cues are not physical objects. They are intentional entities and require a mind.



Figure 10. The use of depth layers. At top a foreground with a backdrop. At bottom foreground, middle ground and backdrop. Each single layer is drawn orthographically, though the scale changes with (average) depth. The layers (or coulisses) are treated as trellises, so deeper ones are seen through the apertures of nearer ones. In the picture plane the layers meet and their planar interrelations are important in design. As Hildebrand says the layers “shake hands in the picture plane”, although staggered in depth. In more sophisticated constructions, one uses objects such as roads that cover large depth ranges continuously and interact with various layers. But Hokusai’s constructions have the advantage of simplicity and speak clearly.

Although very general, the model predicts many observations that can easily be made by informal means. Perhaps, surprisingly, such observations include huge discrepancies between visual awareness and what is referred to as “veridical perception”. Although not proper psychophysics, many huge effects never made it into the textbooks. Even if experienced, many scientists will dismiss the presentations in their awareness. This may be okay from the standpoint that textbooks should only treat topics of proper science, but it is at odds with the requirements of many practical problems. For issues involving human mental processes, one relies on experimental phenomenology, rather than science proper.

Many of the effects mentioned here have been familiar to visual artists for centuries and are widely used, but are essentially unknown — or at least ignored — by vision science. There have hardly been serious attempts at formalization, it remains “how-to” knowledge propagated through workshop practice.

The present paper at least offers an attempt.

ACKNOWLEDGMENT

We acknowledge structural funding from the Flemish government awarded to Johan Wagemans (METH/22/02). The authors are thankful to numerous students and collaborators over half a century of research.

APPENDIX A. VARIOUS FORMAL MATTERS

We succinctly introduce the formal structure that underlies some of the statements in the main text. The topics appear roughly in the order they appear in the text, from an overall perspective rather arbitrarily. We keep the account as simple and short as possible. All the relevant math is readily available from introductory texts.

A.1 Distance and Range

The simplest concept of “remoteness from the self” is *range*. Range is simply the Euclidean distance from the eye.

In contradistinction, “distance” is a formal concept that tends to be not intuitive at first blush. Distance is defined as the separation of any frontoparallel plane from the frontal plane through the eye. It fits the notions of geometrical optics, so distance (not range!) is what is marked on photographic lenses. Distance is the measure of separation from the self that is naturally adapted to the notion of a “frontal observer” (humans, cats and other predators as opposed to sheep and other prey animals).

In our notation the x -coordinate is distance, whereas $\sqrt{x^2 + y^2}$ measures range. They coincide on the principal viewing direction, but — even for a canonical view — may differ very significantly.

A.2 2D and 3D

Because optics relies on straight “rays”, it works out the same in all planes through the eye. Thus a 2D formalism serves as a 3D formalism too. For example, the geometry in the horizontal (XY) and sagittal (XZ) planes is identical. The formalism is easily adjusted by adding or removing the Z and W dimensions and treating the $\{y, v\}$ and $\{z, w\}$ coordinates relations as fully analogous.

A.3 Projective Geometry and Homogeneous Coordinates

We use homogenous coordinates $\{p, q, r\}$ ($r \neq 0$) to represent points $\{x, y\} = \{p, q\}/r$ and (if $r = 0 \wedge (p^2 + q^2) \neq 0$) directions $\{p, q\}$. We set it up such that $\{0, 0, 1\}$ represents the eye, $\{1, 0, 0\}$ the principal viewing direction and the plane $x = 1$ the frontal plane. Points on the frontal plane define visual rays, this is Alberti’s linear perspective (see A.3.2).

A.3.1 The Basic Hyperbolic Involution

The backward-identity $\mathbf{J}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ describes the basic optics. We have $\mathbf{J}_3 \cdot \{x, y, 1\} = \{1, y, x\}$, which defines the hyperbolic involution $\Pi : \{x, y\} \mapsto \{1, y\}/x$.

Note that $\mathbf{J}_3^2 = \mathbf{I}_3$, thus \mathbf{J}_3 is its own inverse. It is a “special square root of the identity”. It has an invariant

point $\{-1, 0\}$ and an invariant line $x = 1$ (frontal plane of the frustum). The point $\{0, 0, 1\}$ (the eye) and the direction $\{1, 0, 0\}$ (the principal vanishing point) are mutually exchanged. Likewise, points $\{x, y\}$ and $\{1, y\}/x$ on lines through $\{-1, 0\}$ are mutually swapped.

A.3.2 Alberti's "Linear Perspective"

A projection of the viewbox on either the hither or yonder pane $\{u, v\} \mapsto v$ yields Alberti's linear perspective. One has $v = y/x$. A projection on the hither pane describes Alberti's notion of the *rete*, conventionally illustrated in textbooks. A projection on the yonder pane reveals the points as the vanishing points defined by the visual rays.

A.3.3 The Hyperfocal Point

Consider a viewport $\{p, q\} = -d/2 < y < +d/2$ in the frontal plane. It defines a quadrangle (the 2D frustum) $PQRS$, where R, S are the *directions* $\{1, \mp d/2\}$. The intersection of the diagonals is $\{2, 0\}$, independent of d . It is known as the "hyperfocal point" (Fig. 3). It is important because it is (by construction) the centre of the frustum and is located exactly "halfway infinity".

A.3.4 Wide- and Narrow-angle Views

For a fixed viewport the viewing distance defines the extent of the field of view $\Phi = 2 \arctan \delta/2$, where δ is the ratio of the viewport width to the viewing distance. The viewing distances shown in Figure A1 left are $1/2, 1$ and 2 . It is about the range used in regular photographic cameras.

Consider a "main topic" that fills the viewport (Fig. A1). We suppose the main topic is to be isotropic (as deep as wide).

For a viewing distance d , it extends into the viewbox from the hither-pane to the depth $1/(1+d)$, here $1/3, 1/2$ and $2/3$. This is very noticeable for common topics.

For wide-angle views, the topic extends beyond the hyperfocal point. Conversely, for narrow-angle views, the topic does not extend to the hyperfocal point, but is concentrated near the hither pane. If the "background" is defined as the region behind the hyperfocal depth ("halfway infinity"), one sees that narrow angle views allow for a "middle ground", whereas wide-angle views do not (Figs. 8 and 9).

A.4 Scenographic Transformations of Window Views

In 2D, the empty region between the eye and the frustum is a triangle, in 3D, a pyramid. If the eye is moved with respect to the viewport this region changes shape. An "equivalent window view" should also transform the empty region appropriately. The immediate geometrical inference is that such equivalences must be due to affine (not projective!) transformations.

This was famously proved by De La Gournerie in the mid nineteenth century. It is often applied to discussions involving the oblique viewing of pictures. Such applications are spurious due to the ontological distinction between window and picture viewing.

A.5 Nil-square Infinitesimals

Consider the imaginary unit ϵ with $\epsilon^2 = 0 \wedge \epsilon \neq 0$. It is definitely small. The number ϵ has no sign, but is different from zero. Thus ϵ is not a real number.

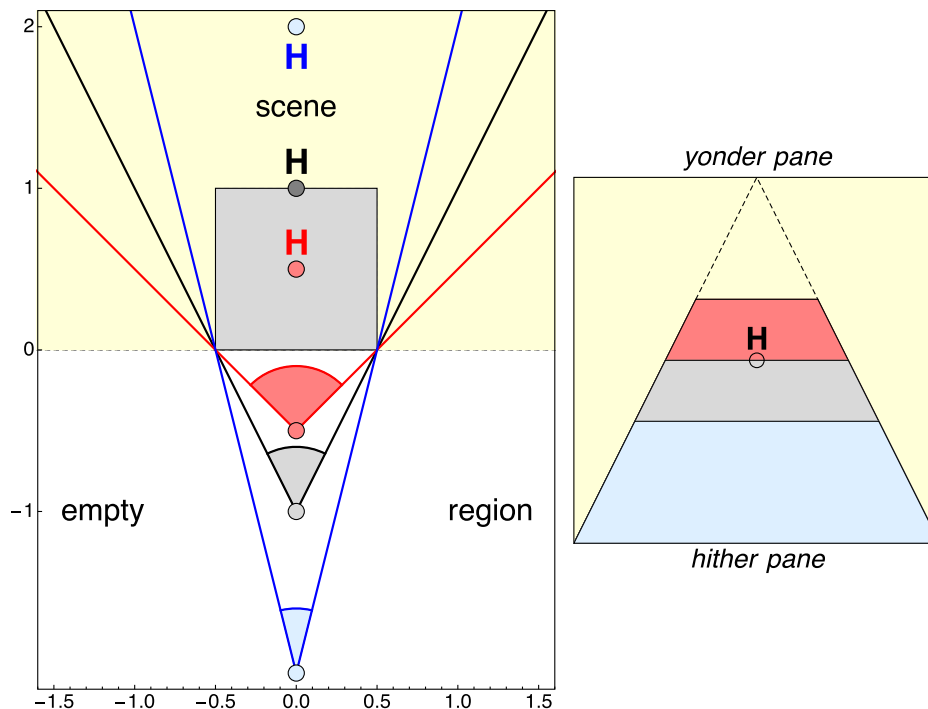


Figure A1. At left we illustrate a change of the viewing distance, keeping the size of the main topic (the square) invariant (section A.3.4). Compare the location of the square relative to the hyperfocal point for the three cases. The viewbox (at right) has unit width in all cases.

The “dual numbers” $z = u + \epsilon v$, $u, v \in \mathbb{R}$ span a non-Euclidean plane with useful properties. One has that $|z_1 - z_2| = u_1 - u_2$, which is zero for points $u_1 = u_2$, that only differ in that $v_1 \neq v_2$. Such points are mutually “parallel.” They may be assigned a special distance $v_1 - v_2$.

All lines of constant u have zero length. These are called “isotropic”. If one treats depth as an isotropic dimension, say ϵd , with $d \in \mathbb{R}$, then the depth may range over all real numbers, yet is contained in a point — a line of length zero. Thus the picture plane can be flat, but still carry a full depth line at each point.

This avoids the inconsistencies that would occur if depth would be treated on a par with the picture plane dimensions.

A.5.1 Mental Transformations

In the picturebox we use coordinates $\{u, v\}$, which we may treat as dual numbers $v + \epsilon u$. Note that transformations of the dual parts have no metrical effects in the picture plane. Specifically, transformations like $u \mapsto u + \tau + \rho v$ are (special) *congruences* (they do not change the special distance), whereas transformations like $u \mapsto \sigma u$ with $\sigma \neq 0$ are (special) *similarities* (these scale the special distance). We denote the group \mathbb{M} of special congruences and similarities the “mental movements” (Fig. 5).

A.6 Ambiguities of “Shape from X”

Many “shape from X” algorithms are bilinear problems of the type $C \cdot E^T = D$, where C is a *Calibration* matrix, E an *estimation* matrix and D an *observation* matrix.

A special solution is immediately obtained through singular values decomposition $D = U \cdot W \cdot V^T$, where W is diagonal. So an “honest” solution is $C = U \cdot \sqrt{W}$ and $E = V \cdot \sqrt{W}$.

But then $C' = C \cdot A$ and $E' = E \cdot \bar{A}$ with $\bar{A} = A^{-1T}$ and $\det A \neq 0$ is a solution too. There is no further constraint on A , thus we obtain a group of ambiguities A . This is typical for many *sfx* problems [23].

Thus the estimated structure (E) remains observable only up to such ambiguities. For “Shape from Shading” one immediately obtains the well known “relief transformation”.

In generic cases the group of ambiguities A coincides with the group M of mental movements (see A.5.1), which is important from a biological perspective.

A.7 The “Abacus Model”

The “Abacus Model” (Fig. 6) essentially treats pictorial space as a fiber bundle. The base space is the mental image of the picture plane. It has (by and large) the structure of the Euclidean plane \mathbb{E}^2 . The fibers are the depth dimension, which is approximately an affine line \mathbb{A}^1 . There is no depth origin, nor something like a natural “unit”. Best one can do — depending on the settings — is to judge ratios of depth differences. Thus a formal model of pictorial space might be $\mathbb{E}^2 \times \mathbb{A}^1$.

Singly isotropic space (see A.5) serves as an effective algebraic model.

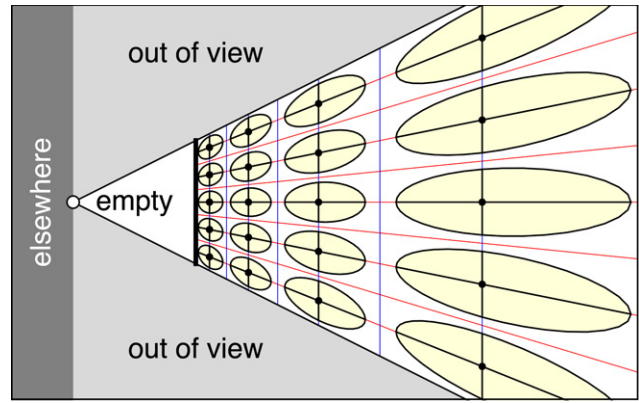


Figure A2. Non-Euclidean metric in the frustum induced by a Euclidean metric in the viewbox. Here we plot equal sized circular disks with diameters in the u and v directions from the viewbox to the frustum.

A.8 The Default Metric

Assume the Euclidean metric in the viewbox. Then this metric inherited from the viewbox induces a non-Euclidean metric in the scene (Figure A2). The metric tensor is

$$G = \begin{bmatrix} 1 + y^2/x^4 & -y/x^3 \\ -y/x^3 & 1/x^2 \end{bmatrix}.$$

All components of the Riemann tensor are zero, so the scene is a flat space in this metric, though evidently not Euclidean. The geodesic equations are readily integrated. The geodesics are straight lines with a projective parameterization.

For instance, frontal geodesics at distance x_0 with $y \in [y_0, y_1]$ are $\{x_0, y_0 + s(y_1 - y_0)\}$ with $s \in [0, 1]$ (like in the Euclidean metric). The principal viewing direction $x \in [1, \infty]$ is a geodesic $\{1/1 - s, 0\}$ with $s \in [0, 1]$ (a nonlinear division, unlike the Euclidean metric).

REFERENCES

- 1 L. Albertazzi, “Experimental phenomenology: An introduction,” in *Handbook of Experimental Phenomenology: Visual Perception of Shape, Space and Appearance*, edited by L. Albertazzi (John Wiley & Sons, Chichester, UK, 2013), pp. 1–37.
- 2 Alberti L. B., *Della Pittura* (Penguin Classics, London, UK, 1972), (In English: On Painting) original 1435.
- 3 A. Ames Jr., “Depth in pictorial art,” *Art Bull.* 8, 5–24 (1925a).
- 4 A. Ames Jr., “The illusion of depth from single pictures,” *J. Opt. Soc. Am.* 10, 137–148 (1925b).
- 5 P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, “The bas-relief ambiguity,” *Int. J. Comput. Vis.* 35, 33–44 (1999).
- 6 G. Berkeley, *An Essay towards a New Theory of Vision* (Bookseller at Skinner-Row, Dublin, Ireland, 1709), Printed for Aaron Rhames, for Jeremy Pepyat.
- 7 E. Claparède, “Stereoscopie monoculaire paradoxale,” *Ann. d’Oculistique* 132, 465–466 (1904).
- 8 R. Cole, *Perspective* (Seeley, Service & Co., London, UK, 1921).
- 9 D. Coxeter, *Introduction to Geometry*, 2nd ed. (John Wiley, New York, NY, USA, 1989).
- 10 M. Denis, “Définition du néo-traditionnisme,” *Art Crit.* 65, 556–558 (1890).
- 11 C. R. Dunn, *Conversations in Paint: A Notebook of Fundamentals* (Workman Publishing Company, New York, NY, USA, 1995).
- 12 C. von Ehrenfels, “Über gestaltqualitäten,” *Vierteljahrsschrift für wissenschaftliche Philosophie* 14, 249–292 (1890).
- 13 Euclid, “The Optics of Euclid,” *J. Opt. Soc. Am.* 35, 357–372 (1945) (translated by H. E. Burton).

- ¹⁴ J.D. Foley, A. van Dam, S.K. Feiner, and J. Hughes, *Computer Graphics: Principles and Practice in C*, 2nd ed. (Addison-Wesley, Reading, MA, 1995).
- ¹⁵ E. H. Gombrich, *Art and Illusion: A Study in the Psychology of Pictorial Representation* (Princeton University Press, Princeton, NJ, 1961).
- ¹⁶ J. De La Gournerie, *Traité de perspective linéaire contenant les tracés pour les tableaux, plans et courbes, les bas-reliefs et les décorations théâtrales, avec une théorie des effets de perspective* (Dalmont et Dunod, Paris, France, 1859).
- ¹⁷ J. Ham, *Drawing Scenery: Landscapes and Seascapes* (Perigee Books, New York, NY, USA, 1972).
- ¹⁸ H. Hesse, "Das Glasperlenspiel. Versuch einer Lebensbeschreibung des Magister Ludi Josef Knecht samt Knechts hinterlassenen Schriften," *Herausgegeben von Hermann Hesse. 2 Bände* (Fretz & Wasmuth, Zürich, Switzerland, 1943).
- ¹⁹ A. von Hildebrand, *Das Problem der Form in der Bildenden Kunst* (Heitz & Mündel, Strassburg, Germany, 1893).
- ²⁰ D. Hoffman, "The interface theory of perception: Natural selection drives true perception to swift extinction," *Object Categorization: Computer and Human Vision Perspectives* (2009), pp. 148–166.
- ²¹ J. J. Koenderink, "Different concepts of "ray" in optics: link between resolving power and radiometry," *Am. J. Phys.* **50**, 1012–1015 (1982).
- ²² J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers, "Surface perception in pictures," *Perception Psychophysics* **52**, 487–496 (1992).
- ²³ J. J. Koenderink and A. J. van Doorn, "The generic bilinear calibration-estimation problem," *Int. J. Comput. Vis.* **23**, 217–234 (1997).
- ²⁴ J. J. Koenderink, A. J. van Doorn, H. de Ridder, and S. Oomes, "Visual rays are parallel," *Perception* **39**, 1163–1171 (2010).
- ²⁵ J. J. Koenderink, A. J. van Doorn, and J. Wagemans, "Depth," *i-Perception* **2**, 541–564 (2011a).
- ²⁶ J. J. Koenderink, "Gestalts and pictorial worlds," *Gestalt Theory* **33**, 289–324 (2011b).
- ²⁷ J. J. Koenderink, "Vision as a user interface," *Proc. SPIE* **7865**, 786504 (2011c).
- ²⁸ J. J. Koenderink and A. J. van Doorn, "Gauge fields in pictorial space," *SIAM J. Imaging Sci.* **5**, 1213–1233 (2012a).
- ²⁹ S. C. Pont, H. T. Nefs, A. J. van Doorn, M. W. A. Wijnjtes, S. F. te Pas, H. de Ridder, and J. J. Koenderink, "Depth in box spaces," *Seeing Perceiving* **25**, 339–349 (2012b).
- ³⁰ J. J. Koenderink, "Ontology of the mirror world," *Gestalt Theory* **37**, 119–140 (2015).
- ³¹ J. J. Koenderink, A. J. van Doorn, B. Pinna, and R. Pepperell, "Facing the spectator," *i-Perception* **7**, 1–29 (2016).
- ³² J. J. Koenderink, "Vision, an optical user interface," *Perception* **48**, 545–601 (2019a).
- ³³ J. J. Koenderink, *Sentience* (Cloodcrans Press, Trajectum, 2019b).
- ³⁴ J. J. Koenderink, *Between Sentience & Sapience* (Cloodcrans Press, Trajectum, 2019c).
- ³⁵ W. Metzger, *Gesetze des Sehens. Herausgegeben von der Senckenbergischen Naturforschenden Gesellschaft zu Frankfurt am Main* (Waldemar Kramer, Frankfurt am Main, Germany, 1953).
- ³⁶ M. H. Pirenne, *Optics, Painting & Photography* (Cambridge University Press, Cambridge, UK, 1970).
- ³⁷ T. Reid, *An Inquiry into the Human Mind*, 3rd edn. (Cadell, London, UK, 1769).
- ³⁸ H. Sachs, *Ebene Isotrope Geometrie* (Vieweg-Verlag, Braunschweig and Wiesbaden, Germany, 1987).
- ³⁹ H. Sachs, *Isotrope Geometrie des Raumes* (Vieweg-Verlag, Braunschweig and Wiesbaden, Germany, 1990).
- ⁴⁰ H. Schlosberg, "Stereoscopic depth from single pictures," *Am. J. Psychol.* **54**, 601–605 (1941).
- ⁴¹ K. Strubecker, "Geometrie in einer isotropen Ebene," *Math. Naturwiss. Unterricht* **15**, 297–306, 343–351 and 385–394 (1962).
- ⁴² J. von Uexküll, *Theoretische Biologie* (Springer, Berlin, 1920).
- ⁴³ J. von Uexküll, *Bedeutungslehre (Bios, Abhandlungen zur theoretischen Biologie und ihrer Geschichte sowie zur Philosophie der organischen Naturwissenschaften. Bd. 10)* (Verlag von J.A. Barth, Leipzig, 1940).
- ⁴⁴ I. M. Yaglom, *Complex Numbers in Geometry* (Springer, New York NY, USA, 1968), Abe Shenitzer (trans.).
- ⁴⁵ I. M. Yaglom, *A Simple Non-Euclidean Geometry and its Physical Basis: An Elementary Account of Galilean Geometry and the Galilean Principle of Relativity* (Springer, New York NY, USA, 1979), E. J. F. Primrose (trans.).