Vision as a User Interface

Jan Koenderink[a,b]

[a] Man Machine Interaction Group, EEMCS, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands;
[b] Laboratory of Experimental Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, 3000 Leuven, Belgium

ABSTRACT

The egg-rolling behavior of the graylag goose is an often quoted example of a fixed-action pattern. The bird will even attempt to roll a brick back to its nest! Despite excellent visual acuity it apparently "takes a brick for an egg." Evolution optimizes utility, not veridicality. Yet textbooks take it for a fact that human vision evolved so as to approach veridical perception. How do humans manage to dodge the laws of evolution? I will show that they don't, but that human vision is an idiosyncratic user interface. By way of an example I consider the case of pictorial perception. Gleaning information from still images is an important human ability and is likely to remain so for the foreseeable future. I will discuss a number of instances of extreme non-veridicality and huge inter-observer variability. Despite their importance in applications (information dissemination, personnel selection, ...) such huge effects have remained undocumented in the literature, although they can be traced to artistic conventions. The reason appears to be that conventional psychophysics—by design—fails to address the qualitative, that is the meaningful, aspects of visual awareness whereas this is the very target of the visual arts.

Keywords: Vision, User interfaces, Ethology, Evolution of vision, Veridicality of perception

1. INTRODUCTION

If there were to be such a thing as the "Standard Model" (SM) of vision*, or perception in general, it would no doubt have to run something close to this:

- that the world exists independently of your vision and that its structure is pretty much as you see it;
- that vision is a process of computation on optical data, that is "inverse optics";
- that the initial stages of vision are implemented through "feature detectors";
- that any ambiguities are resolved through Bayesian estimation;
- that vision results in the "representation of the scene in front of you;"
- that your representation is a nexus of symbols and that these symbols derive their meaning from the fact that they correspond uniquely to physical things;
- & more of these statements deriving the same general mind set ...

The corresponding mind-set is objectivist, representationalist, physicalist and it is anthropocentric.[26] Although commonly associated with the true scientific attitude, it is inconsistent, non-sensical, and at odds with such pillars of modern science as the theory of evolution. Because the SM is so pervasive, I will use much of the introduction to explain above statements.

To believe that the world exists independently of your vision and that its structure is pretty much as you see it, is to ascribe to a unique "God's Eye's" view,[26] and to a special position of the human observer in the

Further author information: Send correspondence to j.j.koenderink@tudelft.nl
*Possible references are the seminal book by the late David Marr[29] and any modern textbook like Steve Palmer's.[32]

Human Vision and Electronic Imaging XVI, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7865, 786504 · © 2011 SPIE-IS&T
CCC code: 0277-786X/11/$18 · doi: 10.1117/12.881671

animal kingdom. The very notion of a God's Eye's view is incompatible with modern science. This was perhaps most clearly explained by Sir Arthur Eddington[8] in his "The Nature of the Physical World" (1928). He uses the example of a physics exam problem involving an elephant sliding down a grassy slope. In the course of his analysis the elephant, the slope, and so forth are eliminated as mere mental paint, and the actual situation is characterized as a set of pointer readings. Such pointer readings are meaningless in the sense of *qualia*. They enter equations and lead to the prediction of possible pointer readings, equally meaningless. One of the latter corresponds to the time it takes the elephant to slide down the slope—to add another touch of mental paint. The mental paint is due to the observer. Different observers may apply different mental paints, this is irrelevant to the physics—that is the "God's Eye's" view if there is any. Whose mental paint is "right," you or your dog's? The SM assumes the anthropocentric attitude, which is of course unscientific.

To identify vision with the running of an inverse optics algorithm is perverse. Algorithms are formal machines that, when started on a formal structure, deliver another structure. For instance, the algorithm "plus" applied to the numbers "2" and "2" yields the number "4." The question "4 what?" does never arise, dollars, galaxies or kisses, it makes no difference. The structures may be understood in terms of actual or potential pointer readings, that happens when a physical structure—an engineered or physiological system say—is described as an algorithmic device. Structure is meaningless unless interpreted—literally endowed with meaning—by some intelligence. The meaning is like a mental paint that doesn't have any influence on the algorithmic device. The same device may well crank out different meanings for different intelligences. A simple example is planar projective geometry,[10] whose axioms are invariant with respect to the interchange of "lines" and "points," thus any theorem has (at least) two distinct meanings. Above, I called the notion of vision as an inverse optics algorithm perverse, because it is a notion that ignores the observer—it is literally inhuman. Such notions apply to zombies, physiological systems without awareness.

The above analysis serves to clarify the categorical distinction between "information" in the sense of structure and in the sense of meaning. Information in the former sense is the topic of Shannon's theory of information,[38] information in the latter sense is the topic of phenomenological psychology. Optical structure available to the eye is often denoted "data," suggesting that a semantics is in place. The same misunderstanding lets the SM populate the visual system with "feature detectors." A feature is a meaningful structural element, thus it cannot be either an input or an output of any algorithmic process. Consider a prototypical feature detector such as an "edge detector," let's say the popular "Canny edge detector."[5] Such an object is actually an algorithm that performs local image processing. Confronted with some image structure it yields a real number. It will do so at any location of the image. This transforms one structure (an image, that is an array of numbers) into another array of numbers.[18] "Edges" occur only in the interpretation of an intelligence that peruses such an array of numbers. Mere computation can transform structure into structure, but cannot endow structure with meaning.

It is usually underestimated how ambiguous optical structures really are.[19] For simplicity, consider a static image, for instance the view of a picture, a photograph say. There exist infinities of virtual scenes that could have produced that identical photograph. Machine vision has only managed to describe certain infinitely dimensional subsets of these.[1] However, even these infinite sets exhaust only subsets of all possibilities of measure zero. The Hollywood film industry has learned to exploit this ambiguity to good advantage. For instance, the "Star Wars" movies were by no way shot in distant galaxies. The idea to deal with such ambiguity by way of Bayesian estimation (an SM notion[17]) is optimistic, to put it mildly. Thus, if vision is proclaimed to result from a "representation of the scene in front of you," then you may well ask "representation of which scene?" Moreover, there isn't any scene in front of you, at least nothing you would know. All there is are pointer readings at best, Democrites' "in reality only atoms and the void" comes close enough.

Notice that photographs are pieces of paper covered with pigments—deposited by an inkjet printer say—in a certain simultaneous arrangement. From the photograph itself you cannot derive whether it is the image of any scene at all. This is up to forensic investigation. It may turn out to be the case that the so called photograph is actually a professional drawing or an overgrowth of tinted fungus, the possibilities cannot be exhausted. To treat the photograph as the "imprint" (a bit like a foot print) of a scene, is to stick your neck out. It is to interpret structure in terms of meaning. Structures themselves are meaningless. This equally applies to the retinal image or any neural activity. That is why neurophysiology will never find the mind,[21] mind is not a physiological structure.

The idea of a "representation" implies the objectivist idea of a God's Eye's view, that there really is something to represent.[26] If there were, then the notion of "veridicality" would make sense. You see veridically if your representation does not conflict with God's Eye's view. Notice that you would need some form of ESP to know God's Eye's view though, so whether your view is veridical is hard to check. You would need both your view and God's Eye's view in order to make the comparison, the notion is inconsistent. Moreover, your view is colored with mental paint, whereas God's Eye's view is colorless, consisting of mere pointer readings. The idea that the representation is a nexus of symbols, somehow—magically—corresponding to real entities is nonsensical. Yet it pervades AI, cognitive science and machine vision. Only "outmoded" approaches like these of the early Gestalt school[30] or Husserl's[14] phenomenological psychology escape this trap.

This concludes some observations on the SM of perception. In this paper I describe alternatives that appear to make more sense, and are closer to biological various notions, including the theory of evolution. After explaining the alternative I discuss various consequences for the theoretical and empirical study of vision, or perception in general.

2. THE LIFE WORLD

Ethology is a fairly recent field of endeavor in biology, its founding fathers Konrad Lorenz, Nico Tinbergen and Karl von Frisch shared the Nobel Prize in physiology or medicine in 1973.[31] Of course there were forerunners, I mention especially Jakob von Uexküll (1864–1944). Some of the insights of the ethologists are of immediate interest to my present analysis. Von Uexküll[41] clearly understood the fact that different agents experience different worlds, he coined the notion of "life-world" (G: *Lebenswelt*). He identified the notion of a God's Eye's view as inconsistent, and the idea that human vision somehow approaches to veridicality as the antropocentrism it is. Examples by Tinbergen and Lorenz were especially striking. Tinbergen[40] identified the "releasers," similar to the Gestalts in human vision, meaningful threads of visual awareness that are immediate and cognitively inpenetrable. A famous example is the red belly of the male stickleback fish. Otherwise quite bizarre (non-realistic) objects worked as a potent "super-pornography" on the female fish. Lorenz[27] discovered the strange "egg rolling" fixed action pattern in the graylag goose. Birds with excellent visual acuity took a brick for an egg! Such cases of perhaps unexpected meanings imposed on almost arbitrary optical structures, clearly show the inapplicability of naive notions of "veridicality."

As von Uexküll already showed such apparently strange mental colorings can lead to excellent results in the game of evolution. Indeed, the evolutionary pressure is by no means towards increased veridicality, but rather towards increased utility. The visual system evolves such as to enable efficacious actions, not towards a best approximation of a God's Eye's view. This is the major, enormously important, insight of ethology. It shows up visual awareness as an *idiosyncratic user interface*, and it offers a causal explanation for this. Since humans are no less animals, one concludes that human visual awareness is an idiosyncratic user interface too.

3. THE GENESIS OF VISUAL AWARENESS

Visual awareness is proto-mind stuff.[†] Here I am thinking of the Gestalts or the classical illusions,[30] which are cognitively impenetrable. Even if you are an expert on a certain visual illusion, you still *see* it, cognition doesn't make it go away. Likewise, a Gestalt like the famous Kanizsa[15] triangle is stubbornly there—in your visual awareness, even though you know that "there is no triangle." Visual awareness is *presentation* in the sense that it *simply happens to you*. Nothing you can do about it, nor are you responsible for the presentations you have. It is much like sneezing in this respect.

The genesis of visual awareness should really be known as "psychogenesis," but is actually known as "microgenesis," the equivalent of the German *Aktualgenese*.[‡] Microgenesis occurs behind the screen so to speak, all you ever now are its outcomes, and there is no way you can influence it, except—but only over appreciable time-spans—by changing yourself. Phenomenologically, microgenesis is a legato–style systolic process, that

[†]Thus awareness is proto-consciousness. Consciousness proper is often defined as self-consciousness, it implies the existence of some concept of self. Mere awareness does not, it is a pure feeling.

[‡]The term psychogenesis being already taken by something irrelevant to the present topic.

arises from the deepest levels of the mind, progresses by progressive articulation, and ends in threads of awareness.[4] Thus awareness is like the outer crust of the microgenetic process, where there is no further possibility of articulation. The moment you become aware, the awareness turns into thought, and is replaced by the next awareness. Thoughts are yours, they are part of what you do, and you are responsible for them. Microgenesis is like biological evolution, involving boundless generation of variety, combined with merciless pruning. In visual awareness the pruning is due to a constant check of the plentiful variety—involving dream-like memories and so forth—against the actual optical structure. This checking and pruning encertains your anchoring in the world. It ensures that your visions are likely to subserve efficacious actions. It has been said that "perception is controlled hallucination," and this is indeed much to the point. At least, if properly understood.

Since visual awareness is spontaneously generated, it is your own construction, and—because of that—meaningful. Awareness *is* meaning. Your awareness grows like a plant, by progressive differentiation and pruning, not at all by way of the assemblage of parts as the SM has it. Meaning derives from wholes, it is not an "emergent property" due to the accumulation of meaningless parts, such would be magic.

3.1 The Sherlock Model

A simple model that serves to understand the use of "controlled hallucination" is what I refer to as the "Sherlock Model." Sherlock Holmes is a fictional, though famous, detective that sprung from the mind of Sir Arthur Conan Doyle[6] (1859-1930). His "method" illustrates the power of creative invention, growth by progressive differentiation, and pruning against physical evidence, that characterizes the microgenetic process. This interplay of induction and deduction also characterizes the "scientific method."

Consider the task that faces a forensic investigator confronted with an obvious scene of a crime, say the dead body of the count with a knife sticking from the chest. What to do, since inverse optics is clearly of no help here? What are the clues? Well, there aren't any. The scene is just structure, and *anything* could be a valid clue. The clueless village policeman at the scene simple collects anything out of the usual, like a cigarette butt in the flower pot, a torn newspaper, and so forth. There is no end to what might be collected. Pure deduction rarely solves a crime. In order to find clues one has to know what to look for. In order what to look for one needs a plot. For instance, the plot might simply be "the butler did it." The origin of the plot is irrelevant. It might be pure fancy, a random thought, or the result of reading too many detective stories.

Given a plot the detective knows what to look for, due to a general background in the way the world is composed. For instance, butlers tend to develop flat feet in the course of their career (so take another look at the floor for signs of marks), they are wont to cover traces of their smoking, and so forth. The upshot is that the plot allows one to look for clues in an organized manner. What if the plot obviously doesn't work out? No problem, one simply tries the next one. Maybe the countess did it! There aren't that many possibilities. One is bound to win any "twenty questions" game against nature. Changing the plot means that the former clues need to be reconsidered (was there lipstick on the cigarette butt?), some discarded (signs of flat feet are irrelevant), others reinterpreted, and that the search for different clues may proceed.

The important point is that the investigator asks questions, interrogates the available structure. Only asking questions is going to generate meaningful data, the meaning being imposed by the questions. A question implies possible answers, this is the meaning. Thus questions both select structure and endow structure with meaning. Questions *probe* the available structure for meaning, and only if the probing meets a *resistance* meaning is generated. Meeting resistance is where awareness is created (Schrödinger's[37] spark of enlightenment).

3.2 Perception as an Idiosyncratic User Interface

Microgenesis functions much like Sherlock Holmes' method of criminal investigation. This means that it is very far from "inverse optics," which is directed world–to–mind. Instead, it works by active probing[§] on the basis of creative "plots," thus it is directed (proto-)mind–to–world. In perception such plots are often known as "situational awareness."[¶] I say "proto-mind," because microgenesis runs in proto-awareness, you become only

[§] This is perhaps Whitehead's "appetition."

[¶] The plots can be likened to "situational awareness," "gist" due to bottom-up processes in the visual system, "frames" in AI, John Searle's "background," and so forth.

aware of the outer crust, where the probing meets resistance that can't be further probed into. The solidified crust appears as "reality,"[4] populated with objects like rocks and elephants. Perhaps perversely, such objects are often regarded as the causes of perceptions whereas they are really its final effects. This misconception is the root of a great many others. The objects in your awareness are figments of your mind, not representations of entities that exist independent of mind. Of course this in no way implies that you would be unfit to act efficaciously when guided by visual awareness.

Because the direction of microgenetic development is (proto-)mind–to–world, visual awareness is meaningful, that is intentional (concept of intentional reference due to Brentano[2]), something that is unthinkable in the world–to–mind, that is inverse optics, conception. Due to the pruning of otherwise creative constructions, the result is likely to subserve efficacious action, but otherwise the ontology of awareness implies idiosyncracy. This might be considered a problem. If you and I happen to stand next to each other, do we experience the same scene? The answer is clearly no. Will we be able to coordinate our actions given such different awarenesses? The answer is clearly yes. Our life worlds are unlikely to be identical, but we both manage to live our lives, meaning that our actions are likely to mesh. Of course this depends upon our being what we are. People from very different cultures will have more different life worlds, and the overlap in the life worlds of you and your dog will be small (though significant).

One way to think of this is to assume that different people will have individually different interfaces[12] to physical and social worlds that are similar for both. They may experience similar features of their life worlds differently, but are likely to be ready for similar actions. It is like you and me were using different interfaces—an Apple computer and a Windows PC say—to a scientific paper we were writing together. I may be surprised, or even puzzled, at what I see when I look at your screen, and vice versa. In perception we can't "look at each other's screens," we simply assume that all people are sufficiently similar for many practical purposes. This often works out fine, though by no means invariably so.

4. CONSEQUENCES FOR THE STUDY OF PERCEPTION

If visual awareness is non-veridical and idiosyncratic one expects to find instances of that in the experimental psychology of perception. However, we also expect some notable exceptions, that is say, cases where generic human observers are likely to agree. Here are examples of the latter:

- measurements of visual abilities that are determined by physiological constraints are likely to agree. Examples are absolute and incremental thresholds for luminance or spectral composition, visual acuity, and so forth;
- measurements that rely on the registration of optically coordinated motor actions are likely to agree. Examples are optically guided reaching actions, and so forth;
- almost any measurement that has been rendered "objective" by minimizing the possible role of awareness, examples include forced choices.

All such methods address the physiology rather than awareness. Awareness exists in the first person account and thus cannot be addressed by fully objective, third person accounts. The study of awareness is squarely in the realm of psychology proper. Because there is a strong tendency in experimental psychology towards the development of objective measures and methods, the influence of awareness is minimized. One result of this trend is a general feeling that awareness is not important, and hardly more than an epiphenomenon. This degrades humans to zombies by turning psychological issues into physiological ones. Some scientists go as far as to say that human speech is really the movement of air molecules.[39] They are obviously "not wrong" in this, but such a view is perhaps hardly useful in psychology proper, nor in real life. The meaning (including emotional overtones and so forth) is clearly what speech really is. This is evidently a subjective view, neither the hard-core physicist nor your dog are going to understand what you mean with such gibberish. Psychology proper cannot be purely objective, for then it turns into physiology and considers humans to be mere meat machines.

Well known instances in which awareness is essential are the so called "visual illusions" and "Gestalt phenomena,"[13] which are often said to be "cognitively impenetrable" which goes to say that they are presentations that simply happen to you, much like the releasers of ethology. These are like objects of a user interface in that they are in no way representations of physical objects. Examples of idiosyncracies are rare, except in cases where observers have different life worlds. Thus speakers of very different languages often have difficulties in perceiving phonemes that belong to the other language.[33] It seems a priori likely that inhabitants of very different ecologies and (for generations!) different life styles, say hunters in the rain forest and farmers on arid mesas, will have different repertoires of Gestalts. Gestalts are most likely an adaptation to ecological statistics.[3,36] This should be a worthwhile research objective.

5. SOME RELEVANT PSYCHOPHYSICAL FINDINGS

I have been "collecting" obvious idiosyncracies of perception for some time. Although there is no mainstream literature on the topic, it proved to be easy enough to find examples galore of extreme inter-individual differences, one even wonders how these could remain undocumented for so long. I give a few examples here, but I am convinced that even a superficial search will reveal many more instances. Such differences are no doubt important in many applications, because typical users are hardly familiar with their physiological make-up, but intimately familiar with their awareness: *for that's what they feel they* ARE*!*

5.1 The Extent of the Visual Field

The field of view of the generic human observer is about a half-space. There are many complicating factors, for instance, the vignetting by brows and cheeks, the eye movements, the binocularity, and so forth. Given the variability in the extent of the *experienced* field of view such complications may be safely ignored though. There are various indications that the "apparent visual field" may be much less than a half-space. The ancient Greek authors specify the visual field as a cone with top angle of exactly a right angle, for reasons that appear irrelevant to us moderns.[7] Kepler,[16] writing in the early seventeenth century, reports—apparently to his amazement—that his apparent visual field was much more restricted than a half-space. Helmholtz,[9] writing at the end of the nineteenth century writes that he knows the visual field to be a half-space, but that it appears to him as a cone with a top angle close to a right angle. Helmholtz also proves a theoretical analysis based on the kinematics of human eye movements that purportedly explains this surprising observation.

Since there were no true data available, I set out to measure the extent of the visual field for a large number of generic human observers.[20] To this end I procured a translucent hemisphere of large diameter, which I fastened behind a fake wall, leaving open a two inches wide hole at the center of the hemisphere. The size of the hole was chosen to allow "key hole vision" with just one eye. The hemisphere was illuminated from behind. On its interior surface I pasted over a thousand black cardboard circular disks. When you put your eye near to the aperture, using head and eye movements to see in every possible direction, you became aware of a luminous, limitless extended surface, covered with a random polka dot pattern. This is what all but one of about eighty observers saw.[‖]

I set a group of naïve observers to the task of drawing a horizontal cross section of the surface they saw through the key hole. Of course there are a great many precautions to take in order to render this task a sensible one. I then measured—among other things—the extent of the apparent visual field from these drawings. The results are surprising. The majority of people apparently experienced a visual field that measures only about ninety degrees across, thus neatly emulating the accounts of the classical Greek authors, and the keen observers Kepler and Helmholtz. A smaller subset came up with very different reports though. One, small group reported very large extent, exceeding that of a half-space. The largest apparent visual field reported in the experiment was two-hundred-and-twenty degrees across! Another, also small group reported visual fields as narrow as twenty degrees across. The actual values may well have been smaller still, since—after all—people have to draw *something*. Apparently this latter group experienced everything right in front of them. This is how Helmholtz describes—to his own surprise—his own awareness, everything appears much more in front of the self than might be expected on the basis of the optics.

So what to make of this? Apparently inverse optics comes not even close in predicting these results. What appears to be the case is that our visual awareness does not associate locations on the retina with directions

[‖]The single exception was aware of a luminous misty space filled with a swarm of black balls.

in the space in front of the corneas. The visual field has a well defined topological structure, usually known as Lotze's local sign.[28] Apparently this has to be augmented with what might be called an "external local sign." The latter is part of the optical user interface. Since the relation to the motor system is not immediately involved, the lack of a "veridical" external local sign is of no consequence to optically guided behavior. All these, mutually diverse, interfaces serve their users equally well.

The bottom line is that *huge* idiosyncratic differences are common in the normal population, but that they have remained unrecorded in the mainstream literature and textbooks.

5.2 The Relative Spatial Attitudes of Objects in your Ken

Suppose you see two mutually congruent objects, both in front of you, and both situated on the ground plane, at equal distances from the eye, though seen in mutually different directions. You are asked to rotate one of the objects about the vertical (by means of some radio-control say) so as to render its spatial attitude parallel to that of the other object. Would you be able to do this veridically? "Veridical" means here "parallel in terms of the Euclidean space in front of the observer." Such a task would be trivial if you had sufficient context. For instance, you can easily judge whereas two cars are parallel to each other if each are parked along the same (straight) curb: you simply judge the attitude of each car separately with respect to the curb. But what if there is no context?

I put this to the test in a simple manner.[22] Wireframe cubes in random attitudes were drawn on a large computer screen. The two renderings differed by a rotation about the vertical in pictorial space. No context was offered. Observers had computer control over the rotation (about the vertical) of one of the cubes. Task was to put it parallel to the other one. Parameter in the experiment was the angular difference of the directions in which the two cubes were presented, this angle could be as large as a hundred-and-twenty degrees.

The result of this experiment was perhaps surprising: all observers committed *huge* errors, thus the settings were not even close to being veridical. Even more intriguing, the amount of the error closely followed the angular separation. Thus, when the angular separation was a right angle, the error also was close to a right angle. What apparently happened here is that the observers did not relate the spatial attitude of the object to the absolute coordinates of the Euclidean space in front of them, but rather related it to the local visual directions. It is not that the observers didn't understand the instructions, we made sure of that. The largest errors amounted to over a hundred degrees, thus the violations of veridicality were by no means minor.

It is easy enough to verify these observations in an informal setting. Have two friends stand at the same distance from you, but in mutually ninety degrees different directions. Let one of them face you, let the other assume an attitude parallel to the former. Notice that the latter person is then necessarily seen in profile view and does not at all appear parallel to the former. Now let both friends face you. Notice that they appear mutually parallel. This is best done in an open field, as interiors offer much conflicting context.

These *huge* mismatches have not been recorded in the textbooks of mainstream vision research. They rarely lead to aberrant behavior, in virtually all cases there is amply sufficient context to ensure that. It seems likely that the effect is known and taken into account by cinematographers, especially if they use panoramic images and have to pose a group, but I am not aware of technical literature from this field.

5.3 The Spatial Attitude of Pictorial Reliefs

Pictorial content is a great source of potential idiosyncrasies in perception because it is not immediately related to optically guided motor behavior. Thus even huge inter-individual mismatches are likely to remain "uncorrected." There is simply no need to synchronize.

When you look into a picture, as opposed to looking at a picture, you often experience what may be called "pictorial space." Pictorial space is often populated with pictorial objects, which are bounded by pictorial surfaces known as "pictorial reliefs." A pictorial relief is always visible, pictorial objects have no backsides. Since the eye is not situated in pictorial space, pictorial objects have no distances to the eye. They do have a "depth," though depth is categorically different from "distance" (to the eye). The distance can be specified by some positive number with the physical dimension of length (feet say). The origin of this distance scale is the eye itself. In contradistinction, depths have to be specified as points on the affine line. That is to say, the coordinate can be any sign, the affine line having no natural origin one measures with respect to some arbitrary point. For instance, if one has a set of N depths (from some experiment say) it is convenient to simply set the average to zero. One then has "relative depths" ($N-1$ degrees of freedom). All depths have to be conceived of as relative. Depth is categorically different from distance, including distances in the picture plane. Thus pictorial space does not have the structure of Euclidean three-space. This is obvious enough when you notice that pictorial objects cannot be rotated about axes parallel to the picture plane, as this would reveal their back sides, an obvious absurdity since the back sides are not in the picture.

A patch of pictorial surface that is not parallel to the picture plane is called "slanted." The slant can be specified as a number, a non-Euclidean angle. The direction of the slant can vary, it may be characterized by the direction of the depth gradient. It is known as the "tilt." The tilt can be specified by a Euclidean angle in the picture plane if one (once and for all) specifies a reference direction and an orientation. The tilt is periodic with period 2π, whereas the slant varies over the real number line. Zero slant indicates parallelity to the picture plane, infinite values specify that the patch is seen "edge on." A sign change of the slant can be compensated by a π change of tilt, thus it is often convenient to normalize surface slants to non-negative values. Here is how one specifies the slant: take two points along the direction of the depth gradient. The slant is the ratio of their depth difference by their separation in the picture plane. This is a dimensionless number since one expresses depth in terms of the dimension used to specify lengths in the picture plane.

If σ denotes the slant of a patch of relief, then the foreshortening φ of the patch is $\varphi = 1/\sqrt{1+\sigma^2}$. Thus the slant can also be expressed in terms of the foreshortening. Since the foreshortening has both an orientation (in the picture plane) and a magnitude, it specifies the spatial attitude (both tilt and slant) of the relief up to a π turn of the tilt, or—equivalently—the sign of the slant. There are various ways to lift this ambiguity, it is not important in the present discussion.

The foreshortening can be conveniently indicated by the shape of the projection in the image plane of a small (really infinitesimal) circle in the relief. This "indicatrix" is an ellipse, the minor axis specifying the direction of the depth gradient (up to a π turn), and the ratio of the minor to the major axis (the foreshortening) the slant (up to sign). One has $\sigma = \sqrt{b^2 - a^2}/a$, where b denotes the major, a the minor axis. This can be used to *measure* the local attitude of pictorial relief: one instructs the observer to draw the indicatrix. This is easily implemented using computer graphics methods and turned into a powerful psychophysical method. The method has a close affinity to established methods of realistic drawing. Measuring the spatial attitude of the relief at a large (hundreds) number of locations allows one to find the relief through integration. One obtains excellent representations of the pictorial relief from sessions that take less than half an hour.

Let $z(x,y)$ denote the relief, where z denotes the depth, and $\{x,y\}$ the picture plane coordinates. Let z_P, z_Q denote the reliefs due to observers P and Q. Then we find that a multiple regression of z_P with $\{x,y,z_Q\}$ of the type

$$z_P(x,y) = \alpha_x x + \alpha_y y + \beta z_Q(x,y) + \gamma,$$

yields coefficients of variation in the 0.9–1.0 range, whereas straight regression of z_P against z_Q may well yield a coefficient of variation not significantly different from zero. That is to say, the coefficients $\alpha_{x,y}$ are appreciable. Notice that the coefficient γ is irrelevant, since the depth domain is affine. The coefficient β implies a dilation or contraction of the total depth range, whereas the coefficients $\alpha_{x,y}$ add an overall gradient to the relief, they denote a non-Euclidean rotation.

We routinely encounter[23–25] depth scalings (parameter γ) up to a factor of four, and non-Euclidean rotations up to two (implying a foreshortening of about 0.45). These are *huge* values. They were not known in psychology, but have been discussed by the German sculptor Adolf Hildebrand[11] at the end of the nineteenth century. Thus observers may experience pictorial reliefs that differ enormously in overall attitude and range.

This is one more example of important inter-individual differences between generic observers that fails to be documented in the standard textbooks.

5.4 The Range of Pictorial Depth

Pictorial points are not necessarily located on the same pictorial relief. If they are not, then their depth relation cannot be probed by the method explained above. Alternative, effective methods are easily forged though. I mention two of them that I recently investigated, one method yielding metrical, the other ordinal depth scales.

It is easy enough to measure ordinal depth relations.[43] One uses the method of pairwise comparison. Suppose one specifies about fifty fiducial points in the picture plane, located on obvious landmarks. If such points are shrewdly located a human observer will automatically locate them in pictorial space. I showed the points two at a time and asked the observer which one appeared the closest. There were somewhat over a thousand orderless point pairs, implying that such a session—involving the judgment of all unordered pairs—could be completed in about an hour. After the session I simply counted how many times each given point was judged to be the closer. This generates a depth ranking order with possible ties. The ranking order again predicts the pairwise rankings, thus one has an objective measure of how well the ranking accounts for the judgments. In practice the ranking orders I obtain for "easy pictures" (say painted landscapes) explain about ninety-nine percent of the judgments. Thus I end up with very consistent depth orders. The scatter found in repeated sessions allows an estimate of the depth resolution. Most observers easily resolve dozens of depth layers.

Metrical depth can be obtained in a variety of ways. One method that I investigated involves exocentric pointing.[44] One superimposes the images of a three-dimensional pointer and a target over the picture. If the locations are appropriate both the pointer and the target will be experienced as objects that have been introduced into the pictorial space. Observers find it an easy task to adjust the rendering of the pointer in such a way that it looks like it actually points at the target in pictorial space. This manipulation involves setting both tilt and slant of the pointer. The tilt is trivial in the sense that it involves only the geometry of the picture plane, but the slant is far from trivial because involves pictorial depth. As it turns out to be the case, it makes a difference whether the observer points from **A** to **B**, or from **B** to **A**. Apparently the pre-geodesics of pictorial space are curved arcs in the representation used here. The unique best fitting parabolic arc allows one to convert the **A** to **B** and **B** to **A** slants into an **AB** depth difference. These depth differences then can be explained by way of a pictorial point configuration. Such an N-point configuration involves $N-1$ degrees of freedom, because the absolute depth origin is indeterminate. These $N-1$ degrees of freedom thus should explain $N(N-1)$ observed slants, or—more reasonably—$N(N-1)/2$ best fitting parabolic arcs. For $N > 2$ the number $N(N-1)/2$ is larger than $N-1$, for $N \gg 1$ even *much* larger. Whether the point configuration really "explains" the data can be determined from the variability of repeated measurements.

I find that all observers yield very consistent results, both in the ordinal and in the metrical tasks.

In the metrical tasks I find huge differences in the overall depth ranges between observers, factors of three or more being common. On the other hand, observers turn out to yield equivalent results modulo transformations of the type

$$z_P(x,y) = \alpha_x x + \alpha_y y + \beta z_Q(x,y) + \gamma,$$

already discussed above.

On the other hand, I find that the depth resolution, the number of distinct depth layers as determined via the ordinal method, are very similar for all observers. Even observers with a remarkably shallow depth range distinguish as many depth layers as observers that apparently experience an enormous depth range.

These findings show that the depth range has nothing to do with resolution. It is more like a "mental paint." Apparently the depth calibration of pictorial depth is almost fully idiosyncratic.

These huge differences in visual awareness between apparently generic visual observers is—once again—not recorded in the standard textbooks at all. It may well be expected to lead to significant differences in the appreciation of pictures (photographs, printed matter, computer screens) though.

5.5 The So-called "Deformations" of Linear Perspective

One often hears mention of the "deformation" induced by wide-angle lenses. Technically this is nonsense: modern wide-angle lenses generally have only very minor deviations from true linear perspective. One has to be an expert to notice these. The above mentioned impression is shared by most observers though. A photograph taken with an extreme tele lens tends to evoke the impression of a squashed space, populated with flattened out objects, whereas a photograph taken with an extreme wide-angle lens tends to produce a space that is dilated in depth, populated with deformed objects.

Mainstream wisdom has that these impressions will go away when the eye is put at the correct perspective center.[34] This is extremely awkward though, and a condition that is certainly not convenient, or even desirable, in daily life settings. At normal reading distance the tele shot would have the size of a postage stamp, and the wide-angle shot the size of a full newspaper page. In either case the picture would be hard to look at. Anyway, if you follow common wisdom, you will notice that the recipe doesn't work out: the so-called deformations stubbornly stay. In order to make them go away you have to take special measures, such as using a peep-hole and removing context.

For any object, there is a certain way it should be rendered for any given viewing distance. This is simply linear perspective. If it works, observers should be able to identify the "correct" picture among a set of contenders for any given viewing distance. This is easily turned into a simple experiment.[35] I drew a wireframe cube on a computer screen and gave observers remote control over the perspective of the rendering. Their task was to set the rendering such that the picture looked like the picture of a cube. Parameters were the size of the picture and the viewing distance. Both were varied over a huge range, the angular size of the cube varied from a few degrees to about a hundred-and-twenty degrees of visual angle.

The results of the experiment are clear-cut. All observers adjusted the same perspective in all conditions, instead of varying it appropriately. Apparently observers couldn't care less about mainstream wisdom. They always adjusted the rendering to that of a cube seen from a distance of about two times its edge length.

Thus observers apply a template of "how a cube should look" instead of take perspective into account. This is the case even if they know the viewing distance and picture size from the context. It is a clear example of a non-veridical user interface. This oddity evidently serves to explain the "deformations" commonly reported for wide-angle and tele photographs. One notices large diffferences from the templates, that are the *familar pictures of familiar shapes.*

Here one has a good example of the importance of the optical user interface in applications. It is well known that painters have applied various "corrections" to "perfect linear perspective" for ages, with the obvious intention to make things *look right.* Painters know from experience about optical user interfaces and apply such knowledge in their praxis.

6. APPLICATIONS

Are the very significant idiosyncrasies of visual awareness of any importance in applications? In my view, most certainly! But I have to make a caveat here: the possible importance of these idiosyncrasies is restricted to *visual experience* and is irrelevant with respect to visuomotor abilities or discriminations that do not involve the qualitative aspects of visual awareness.

Of course there remain many situations in which the qualitative aspects of visual awareness are crucial. For instance, this holds for virtually any topic that involves *user experience.*

User experience maybe seen as interesting, but perhaps epiphenomenal from a pragmatic viewpoint. This is a limited view though, because user experience certainly enters cogitation, both on a cognitive and on a pre-cognitive level. This goes down as deep as certain emotional states of which the user may be consciously unaware.

If the qualitative aspects of visual awareness are due to the fact that perception is really an idiosyncratic user interface and that perception does not strife to approach "veridicality," that is to say, some direct connection to an objective, user independent reality (a God's Eye's view), then this should have a major impact on the design of user interactions with anything—animals, fellow humans, or machines.

The industry has learned that it pays to differentiate its products into ones that particularly address males, females, youngsters, elderly, and so forth. It still has to understand that generic members of such groups may mutually differ in ways that make the conventional group borders seem almost irrelevant. The good news is that it is possible to quantify such differences, and that it seems likely that an entirely novel distinction of types is possible.

Of course this implies a major effort that goes significantly beyond the mainstream convictions of experimental psychology. Finding ways to differentiate the population into more or less homogeneous groups with regard to perceptual awareness (of course in various modalities), connected with situational awareness and emotional response, is certain to have numerous social applications not to mention business and industry.

7. CONCLUSIONS

I have discussed the nature of perceptual awareness and argued that it reveals vision to be an idiosyncratic user interface. The perceptual systems evolve so as to optimize utility, that is efficacious action, rather than any approach to "veridicality." Psychology has chosen to assume an extremely anthropocentric position, combined with an objectivist belief in a perceiver independent God's Eye's view. Biology, especially ethology, teaches differently. We should take the understanding that *homo sapiens* is one genus of the animal kingdom seriously.

Once we do, it becomes obvious that optical awareness is an idiosyncratic user interface. It does not reveal the world "as it is," but rather *shields* the user from the sheer infinite structural complexity it has to deal with.[12] The optical world we experience is of our own** construction, that is why the world—which is *our* world—makes sense to us.[42] The qualities, or meaning, we perceive are nothing but our own mental paint. The world "as it is" is nothing but meaningless pointer readings.

Notice that this implies that perception is direct. What you are aware of is reality itself, the world. However, this reality in terms of its qualitative, meaningful aspects, is *your* reality. You *are* reality. There is nothing worth knowing behind it, that is to say, it is not that your mind does its utmost to represent something even more real than what you are aware of right now. What you see is what there is. Of course that goes not to say that my reality is yours, or that you couldn't see more if you tried. There is no single truth about the world, your dog and you are both right.

This has to be sharply distinguished from vision as "optically guided behavior." Most of the visual system is involved in that, most of the time. But it happens outside our immediate awareness, in this description of vision we are zombies. The same is true for much of experimental psychology, which has become something much like "non-invasive physiology." Notice that I'm not saying that this is not important. It is. A keen understanding of physiology—the human as zombie—is just as important as medicine, and so forth. But ignoring awareness comes at a price. The gain of ignoring awareness is objectivity, but the loss is subjectivity, that is to say quality and meaning. The heritage of the early Gestalt schools and of phenomenology is almost forgotten, but if we want to forge a science of quality and meaning it will have to be reanimated.

The industry has mined the physiological approach and used it in the design of audio and visual appliances for instance. Right now industry is more interested in the experiential side, as the physiological side has already been mined. Unfortunately, science has little to offer because it is almost exclusively focussed on so-called "objective" approaches. It is a major challenge to forge a science of the subjective, that is to say, of quality and meaning in awareness. By way of a number of examples I have tried to indicate that this is indeed possible.

ACKNOWLEDGMENTS

I could not have conducted this research on my own. I am grateful to many friends and colleagues, I especially mention Andrea van Doorn, Johan Wagemans, Stijn Oomes, Huib de Ridder, Sylvia Pont, Harold Nefs, Maarten Wijntjes, Astrid Kappers and James Todd. Of course they don't necessarily share my interpretation of the empirical data.

REFERENCES

[1] Belhumeur, P. N., Kriegman, D. and Yuille, A., "The Bas–Relief Ambiguity," International Journal of Computer Vision 35, 33–44, (1999).

[2] Brentano, F., [Psychologie vom empirischen Standpunk], Duncker & Humblot, Berlin, (1874).

[3] Brunswik, E., "Representative Design and Probabilistic Theory in a Functional Psychology," Psychological Review 62, 193–217, (1955).

[4] Brown, J. W., [Self-Embodying Mind: Process, Brain Dynamics and the Conscious Present], Barytown Ltd., Barytown NY, (2002).

[5] Canny, J., "A Computational Approach To Edge Detection," IEEE Trans. Pattern Analysis and Machine Intelligence 8(6), 679–698, (1986).

**Here I include the evolutionary history of the genus of course.

[6] Conan Doyle, Sir Arthur, [A Study in Scarlet], (the first Sherlock Holmes story) appeared in 1882. It is available on the Internet: http://sherlock-holmes.classic-literature.co.uk/a-study-in-scarlet. (1882).
[7] Doesschate, G. ten, "Oxford and the revival of optics in the thirteenth century," Vision Research 1, 313–342, (1962).
[8] Eddington, A., [The nature of the physical world], MacMillan, London, (1928).
[9] Helmholtz, H. von, [Handbuch der Physiologischen Optik], Voss, Leipzig, (1860).
[10] Hilbert, D. and Cohn–Vossen, S, [Geometry and the imagination], Chelsea, New York, (1952).
[11] Hildebrand, A. von, [Das Problem der Form in der bildenden Kunst], (1901).
[12] Hoffman, D., "The interface theory of perception: Natural selection drives true perception to swift extinction." in: [Object categorization: Computer and human vision perspectives], Eds. Dickinson, S., Tarr, M., Leonardis, A. and Schiele, B, Cambridge University Press, Cambridge UK, 148–165, (2009).
[13] Hoffman, D. D., [Visual Intelligence, How We Create What We See], W. W. Norton & Company, Inc, (1998).
[14] Husserl, E., [Logische Untersuchungen. Erster Teil: Prolegomena zur reinen Logik], (1900).
[15] Kanizsa, G., "Margini quasi-percettivi in campi con stimolazione omogenea," Rivista di Psicologia 49(1), 7–30, (1955).
[16] Kepler, J., [Johannes Kepler New Astronomy, trans. W. Donahue], Cambridge University Press, (1993).
[17] Knill, D. C. and Richards, W., [Perception as Bayesian Inference], Cambridge Uni- versity Press, Cambridge, UK, (1996).
[18] Koenderink, J. J., "The brain a geometry engine," Psychological Res. 52, 22127, (1990).
[19] Koenderink, J. J., "Vision & Information," In: Albertazzi, L., [Information], to appear with MIT press, (2011).
[20] Koenderink, J. J., Doorn, A. J. van, and Todd, J. T., "Wide distribution of external local sign in the normal population," Psychological Research 73, 14–22, (2009).
[21] Koenderink, J. J., "Guest editorial: Brain scanning and the single mind," Perception 28, 1181–1184, (1999).
[22] Koenderink, J., Doorn, A. van, Ridder, H. de and Oomes, S., "Visual rays are parallel.," Perception 39, 1163–1171, (2010).
[23] Koenderink, J. J., Doorn, A. J. van, and Kappers, A. M. L., "Surface perception in pictures," Perception & Psychophysics 52, 487–496, (1992).
[24] Koenderink, J. J., and Doorn, A. J. van, "Pictorial space." In: [Looking into pictures: an interdisciplinary approach to pictorial space], Eds. Hecht, H, Schwartz, R. and Atherton, M., MIT Press, Cambridge, MA, Chapter 12, (2003).
[25] Koenderink, J. J., Doorn, A. J. van, Kappers, A. M. L. and Todd, J. T., "Ambiguity and the 'mental eye' in pictorial relief," Perception 30(4), 431–48, (2001).
[26] Lakoff, G., [Women, Fire, and Dangerous Things: What Categories Reveal About the Mind], University of Chicago Press, (1987).
[27] Lorenz, K., [Die Rückseite des Spiegels: Versuch einer Naturgeschichte des menschlichen Erkennens], Piper Verlag, München, (1973).
[28] Lotze, R. H., [Medicinische Psychologie oder Physiologie der Seele], Weidmansche Buchhandlung, Leipzig, (1852).
[29] Marr, D., [Vision: A Computational Investigation into the Human Representation and Processing of Visual Information], Freeman, NewYork, (1982).
[30] Metzger, W., [Gesetze des Sehens], Waldemar Kramer, (1975).
[31] Nobel Prize in Physiology or Medicine for 1973, Karolinska Instituted, Press release http://nobelprize.org/nobel_prizes/medicine/laureates/1973/press.html. (1973).
[32] Palmer, S. E., [Vision Science: Photons to Phenomenology], MIT Press, Cambridge, Mass., (1999).
[33] Pinker, S., [The language instinct], MIT Press, (1994).
[34] Pirenne, M. H., [Optics, painting & photography], Cambridge University Press, London, (1970).
[35] Pont, S., Nefs, H., Doorn, A. van, Wijntjes, M. and Koenderink, J., "When does a Cube *Look* Like a Cube?," Seeing and Perceiving –, –, (submitted).

[36] Ren, X., Fowlkes, C. C. and Malik, J., "Learning Probabilistic Models for Contour Completion in Natural Images," International Journal of Computer Vision 77, 47–63, (2008).

[37] Schrödinger, E., [What is life?], Cambridge Un. Pr., (1944).

[38] Shannon, C. E., "A Mathematical Theory of Communication, Part I," Bell Systems Technical Journal 27, 379–42, (1948).

[39] Skinner, B. F., "The Operational Analysis of Psychological Terms," Psychological Review 52, 270–277, 290–294, (1945).

[40] Tinbergen, N., [The study of instinct], Clarendon Press, Oxford, (1951).

[41] Uexküll, J. von, [Umwelt und Innenwelt der Tiere], Springer, (1909).

[42] Vico, G., [The First New Science, edited and translated by Leon Pompa, (Orig. Scienza Nuova, 1725)], Cambridge University Press, Cambridge, UK, (2002).

[43] Wagemans, J., Doorn, A. van and Koenderink, J., "Depth order in pictures," iPerception –, –, (submitted).

[44] Wagemans, J., Doorn, A. van and Koenderink, J., "Measuring 3D Point Configurations in Pictorial Space," iPerception –, –, (submitted).