Directing the Mental Eye in Pictorial Perception

Jan J. Koenderink[a], Andrea J. van Doorn[a], Astrid M. L. Kappers[a], and James T. Todd[b]

[a]Universiteit Utrecht, PO box 80000, 3508 TA Utrecht, The Netherlands
[b]The Ohio State University, Department of Psychology, 142 Townshend Hall,
Columbus OH 43210, U.S.

ABSTRACT

The optical structure sampled by the human observer is insufficient to determine the structure of a scene. The equivalence class of scenes that lead to the same optical structure can be worked out precisely for specific "cues" (shading, texture, ...). If the "observed scene" is a member of the correct class, the observation must be considered "veridical", *even if the observed scene differs from the actual one.* In many cases it is impossible to indicate the equivalence classes, and it therefore must remain undecided whether observations that deviate from physical reality should be denoted "veridical" or not. For observations on the basis of images, such as straight photographs of physical scenes, the equivalence classes are unknown, but are certainly large. This case is especially important in the design of computer interfaces where a scene is being presented to the user as an image. We find that observations differ appreciably according to the precise task. The observer uses the freedom resulting from the iconic underdetermination to choose some idiosyncratic perspective by directing the "mind's eye". This can be demonstrated with simple means. Stretchings of depth by a few hundred percent and changes in viewing direction (not rotations, but shears) of tens of degrees are quite common.

Keywords: psychophysics, pictorial relief, picture perception, ambiguity, visual cues, depth cues

1. AMBIGUITY AND VERIDICALITY

1.1. Scenes, Images and Perceptions

We consider the visual perception of scenes, either in an actual setting, or in case of photographs of actual scenes. The scenes we consider typically contain generic opaque objects (pieces of abstract marble sculpture say) against simple backgrounds, illuminated according to conventional studio techniques (see figure 1). Photographs are "straight" ones. The psychophysical data presented have been obtained on such photographs of scenes. The major deviation from conventional psychophysics is that the stimuli are complicated and only partly controlled. The photographs differ significantly from computer graphics in that the full bouquet of physical effects (surface BRDF's, multiple scattering, vignetting of extended sources and so forth) influences the result in an exact manner.[2]

Thus we consider the sequence *scene*, *image* and *perception* (a psychophysical response). A *scene* is characterized via geometrical structure, material properties, and a light field. The light field originates from (usually multiple and extended) primary sources, but is conditioned by multiple scattering in the scene and the turbidity of the medium. The *image* derives from the scene through an imaging process, largely determined by the vantage point and viewing direction. The *perception* derives from the image as well as from the observer's knowledge of and expertise in "ecological optics" and the psychophysical method used to "measure" (perhaps better: operationally define) the perception.

Scenes and light fields are well understood.[2] Standard physics suffices to describe scenes in all the detail that one deems necessary. Notice that this is actually more a matter of principle than of practice though. For instance, the full BRDF (Bidirectional Reflection Distribution Function) of preciously few materials is known in detail.[3] In practice our knowledge of the scene will often have to remain fragmentary.

Further author information: (Send correspondence to Jan J. Koenderink)
J.J.K.: E-mail: j.j.koenderink@phys.uu.nl
A.J.vD.: E-mail: a.j.vandoorn@phys.uu.nl
A.M.L.K.: E-mail: a.m.l.kappers@phys.uu.nl
J.T.T.: E-mail: jtodd@magnus.acs.ohio-state.edu

In *Human Vision and Electronic Imaging V*, Bernice E. Rogowitz, Thrasyvoulos N. Pappas, Editors,
Proceedings of SPIE Vol. 3959 (2000) • 0277-786X/00/$15.00

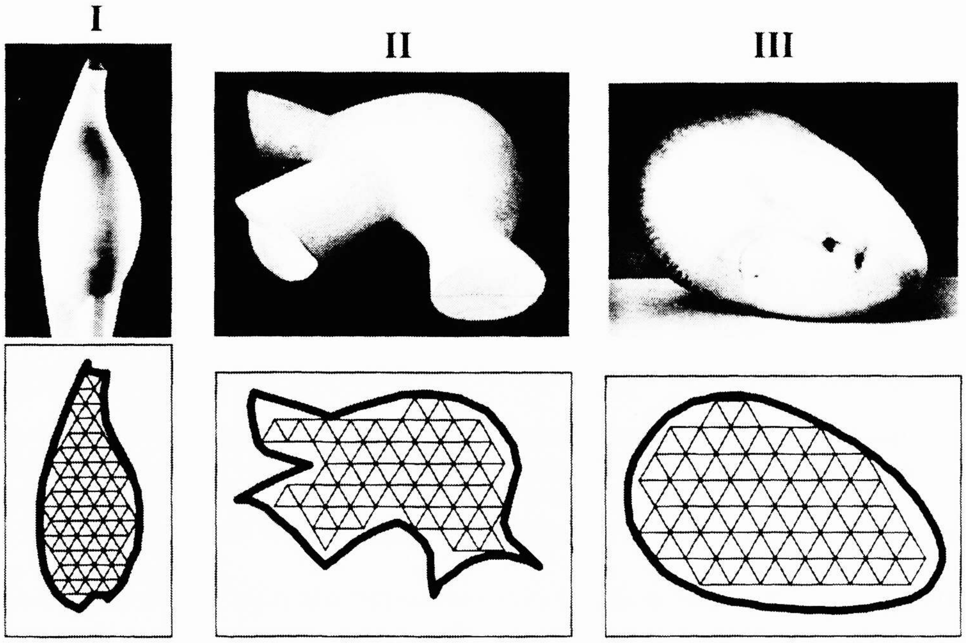

Figure 1. *Upper row: Stimuli used in the experiment. We refer to these as stimulus I, II and III. They are photographs of sculptures by Constantin Brancusi: — I: "Maiastra" (1915–c.1930), Geist[1] catalog 100; — II: "The turtle" (c.1943), Geist[1] catalog 229; — III: "Sleeping muse" (1909–10), Geist[1] catalog 71. Lower row: The triangulations of the photographs used in the experiments (unknown to the observers who only get to see the images).*

The imaging process is also well understood. This is simply *optics*. Given the scene we can in principle predict any image of it. In practice we don't have to of course, we simply take photographs. The fact that we understand the imaging process is important for other reasons though. It is important in order to understand to what extent images and scenes are *related*. For instance, a given scene will yield a certain image for certain camera parameters. But to what extent does a given image specify a specific scene? We will discuss this issue at length.

When we consider the perception of actual scenes (without the intermediary of a photograph) we may consider the *retinal images* as intermediaries. For the stationary observer, monocularly looking at a static scene, the situation is not essentially different from the case where a photograph is introduced in the chain. In this paper we concentrate on pictorial perception though.

1.2. Perceiving the Structure of Scenes: Cues

Scenes are three dimensional distributions of matter whereas images are only two dimensional distributions of gray tones (the introduction of color doesn't change things in any essential way, we'll ignore it here). Thus images necessarily *underdetermine* scenes. For the remarkable bishop Berkeley[4] this was reason to draw the conclusion that it is impossible to *see* scenes. We *deduce* (representations of) scenes and we may be *wrong* in our deductions. Elements of scenes are only known by their marks, not directly.

These "cues" (according to Berkeley) are entirely *arbitrary* associations. For instance, suppose I look someone in the face and I notice a shift of the radiant spectral power density towards the longer wavelength region. Then I "see shame or anger in the face". Why? Because of a learned, arbitrary association. Nowadays many of the cues are well understood in terms of their causal structure.[5] Even Berkeley understood many of them. For instance, when two persons are seen in an image and their sizes in the image are quite different, we tend to see the person with the larger size in the image as nearer. There exists a trivial geometrical explanation based upon Euclid's optics.[6] Despite this causal "explanation" Berkeley insists that the association is an essentially *arbitrary* one. From the perspective of modern biology he must be right.[7] This doesn't at all detract from the fact that we understand why the cue works. It is like understanding the shape of a fish. We understand the torpedo shape from our understanding of hydrodynamics, but the fish cannot be suspected of any such knowledge. Its body was shaped through the random process of evolution. The fact that it has a torpedo shaped body just happened to the fish.

The *cues* of visual perception are understood as *theories* in computer vision. We understand various perspective and photometric cues (shape from shading, and so forth) in intricate formal detail. Most humans don't understand such arcane theories, yet it is obvious from their optically guided behavior that they are quite adept in using them. Even many animals must have sophisticated scene perception. It is indeed much like the torpedo shape of many fishes, we—as scientists—understand why observers are successful in adapting certain visual strategies, the observers themselves (unknowingly) follow the Berkelian pattern. They act the way they do either because they have no other option (phylogenetic learning) or because of a lifelong history of uncontradicted experience (ontogenetic learning).

1.3. Ambiguity

Perceptions from images arise from the exploitation of cues (section 1.2). The cues are read into the image structure by the observer. Much is due to the observer's expertise. An observer able to use "shape from shading" will identify certain gradients in image gray tone as due to a curvature of some diffusely scattering surface with respect to some direction of irradiation.[8,9] It is not at all necessary that this gradient originally arose the way the observer fancies. For instance, ancient marble statues often collect layers of dark dust on their upper surfaces. Many observers spontaneously "see" these sculptures as irradiated from below. They hallucinate a light field.

Suppose one is right in identifying a certain gray tone gradient as due to shading of the kind dear to authors of "shape from shading" algorithms. Does this allow one to infer the shape? Only partially so. It is well known that shape from shading solutions are not unique, but allow for a family of ambiguity transformations.[10–12] What we mean is that there exists a family of (virtual) scenes that agree in that they all would have produced the same photograph (here we reckon the camera as part of the description of the scene). All these virtual scenes are related by transformations that affect the geometry, the light field as well as the surface albedos. These transformations form a group.

Similar results are obtained for the other cues (shape from motion, shape from disparity, and so forth). One guesses (there are as yet no formal results) that combinations of cues will decrease the ambiguity (of course) but in general will never remove it entirely. Ambiguity will remain unless the image equals the scene. Ambiguity is a fact of life. One simply has to live with it. Theories are not up to a complete description of the ambiguity in realistic cases. We only have formal results for very specific, reduced circumstances.

All (virtual scenes) that would lead to the image will be denoted "metameric scenes" in this paper. The actual scene that produced the image will be referred to as the "fiducial scene". It is important to notice that the fiducial scene is only known to the photographer. It is in no way implied by the image, it is only one metameric scene among many. This is obvious when you think of the case of two distinct scenes and two photographers, one for each scene. It is conceivable that the resulting photographs are identical. Clearly the photographs can't imply the scenes then.

The class of metameric scenes is very wide. Given a photograph it includes (among infinitely many more)

- — the fiducial scene;
- — the photograph (a sheet of paper with a certain simultaneous order of pigments);
- — a peculiar stellar constellation viewed from such a point that the radiance sampled by the eye as a function of direction equals that in case the eye views the photograph.

In the latter case we may take the actual stars as very dispersed, for instance no two of them nearer than a few light years apart say. The reader will probably agree that the first two possibilities are somewhat more likely interpretations of the optical structure available to the observer than the third one. (Which is much like the Ames[13] "chair demo".) Indeed, many of the metameric scenes will be freaks in the sense of being very unlikely in an ecological sense. However, there will no doubt exist infinitely many metameric scenes that are ecologically quite reasonable.

1.4. Veridicality

The issue of veridicality is generally considered to be a trivial (though important) one, both in psychophysics and in computer vision. One simply compares the result (percept or result of calculation) with the (fiducial) scene and judges to what extent the two are numerically identical.

This naive concept will not do though. The ambiguity discussed above (section 1.3) implies that many (typically infinitely many) different scenes could have produced the same image. If the percept is not like the fiducial scene but like a metameric scene, can one say that the percept is "not veridical"? We think not. For only the photographer knows the fiducial scene, again, it is not implied by the image. Either we have to say that "veridical perception" is meaningless or we have to say that the percept is veridical if it is like *any* particular metameric scene. The latter definition seems preferable over the former in that it at least allows us to carry on with the investigation.

This problem is a familiar one in the setting of color vision. Any percept of "object color" implies a spectral albedo of an object and a spectral radiant power density of a source that are both highly ambiguous. There has been hot debate over whether the perception of color can *ever* be said to be veridical.[14] One camp holds that color is mere "mental paint", another that one "sees by wavelength". The former opinion is sterile since it places color vision outside science. The latter is mistaken because it denies the ambiguity. Here we see that a quite similar problem pervades the perception of scenes (involving geometry, material properties and light fields). It should make a fruitful field of endeavour to philosophers.

If a percept is said to be veridical if it is metameric with the fiducial scene we get into a number of hairy problems. For instance, at this moment in its development science is not up to the task of verifying whether metamery applies or not. Moreover, if we only have an image we don't know the fiducial scene to begin with. In real life we typically don't know the fiducial scene if we view a photograph (say in a magazine) of course. This is more a problem of practice than of principle though. If the percept indicates a physical scene (or some probability distribution of the space of possible scenes), we can simply apply our knowledge of optics to calculate the image and compare it to the actual image. If they are quantitatively sufficiently similar the percept is one of a metameric scene. The computation[15] essentially is a mere matter of "ray tracing". Thus the definition makes solid operational sense.

1.5. The Observer's Share

Images clearly underdetermine scenes in the sense that infinitely many percepts could equally well be regarded as "veridical" (section 1.4). In psychophysical tasks the observer is typically forced to come up with a unique "solution" though. This can only be done if the observer sticks the neck out and ventures an opinion as to which metamer is most likely to be the fiducial scene. The observer resolves the iconic ambiguity actively, we denote this as "the observer's share" in the percept. Thus we regard the percept as partly causally connected with the structure of the image, partly due to the whim (actually understanding of ecological probabilities) of the observer.

There are many problems connected with these notions. For instance, if the observer is not forced to come up with a unique answer, is the percept (what's in the head when the observer looks at the image) a unique scene or a *set of metameric scenes*? In the latter case the observer might be said to suspend judgment. When the observer has to act it will often be necessary to arrive at a unique interpretation, but if no overt action is required it is conceivable that judgment is suspended. Indeed, many observations seem to indicate that the latter condition applies to human perception. This may be called the "*multiple visual worlds*" hypothesis. It is orthogonal to the conventional interpretation (typically not even mentioned) that perceptions are invariably singular or determined.

The observer's share is actually much broader than implied above. For instance, it is up to the observer to interpret a gradient of gray tone as due to the irradiation of a curved, diffusely scattering surface or not. If the observer does so it is of the nature of an assertion. It is an action of the observer, it is not in the least implied by the image. That is why the fiducial scene and the photograph as a planar object covered with pigments in a certain pattern are two common perceptions due to the same stimulus. This is why animals have no "pictorial spaces" like we do. Animals are no less adept in using most cues than we are. They simply don't apply this expertise to images the way we do.

It has been said that "perception is controlled hallucination". This is at least true in the sense that it is up to the observer to interpret image structure in terms of cues. It also indicates that observers may well differ in the sense that one person's cues may be another person's noise (texture, pattern, ...). The "image" in the visual front end (optic nerve activity say) is ultimate *reality*, whereas the percept is the result of the *constructive imagination* constrained by this image and by the observer's optical expertise and knowledge. The image in the front end is *true* (or "real") because it happens to the observer, like a footprint happens to the beach. It is strictly meaningless (in the sense of being pre–categorical and thus unutterable) though. Perceptions are not automatically true (and can indeed be illusionary in the naive sense) because they are necessarily *constructions.* The "footprint" on the beach might be the result of artistic expression or the chance result of natural forces, like a face in the clouds. Perceptions are by their very nature (because universals, thus utterable) transcendental illusions.

2. PSYCHOPHYSICS AND PICTORIAL RELIEF

2.1. Operationalization of Pictorial Relief

There are many ways to operationalize "pictorial relief". When an observer looks "into" a picture, a "pictorial space" appears and the vision is stopped by "pictorial surfaces". The shape of these pictorial surfaces is known as "pictorial relief". Although the surfaces appear at indefinite depth, they are perceived as having well determined surface orientation and curvature. Moreover, the depth difference of two points on such a surface make perceptual sense. Thus one might attempt to quantify pictorial relief by way of (two point) depth differences, depth gradients (or surface attitudes), or curvatures.[16]

An operationalization of depth differences[17] is simply to indicate point pairs in the image plane and let the observer judge (using a forced choice paradigm) which one is closer. Repeating this for many point pairs allows one to construct pictorial relief.

One good method to measure surface attitude[18–21] is to superimpose a wireframe ellipse over the point to be sampled and let the observer judge whether the ellipse appears as "a circle painted upon the surface". From the orientation and eccentricity of the ellipse one immediately deduces the surface orientation. Integration yields the pictorial relief.

An easy method to quantify curvature and curvature variation is to let the observer draw cross sections in planes spanned by the depth direction and some fiducial line on the surface (see figure 2). Repeating this for many cross sections again allows one to construct the pictorial relief. We show some results obtained with this method as illustration of the relevant phenomena in this paper.

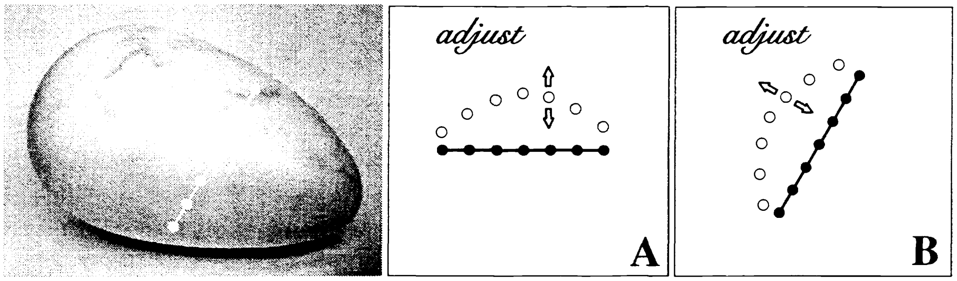

Figure 2. *Method of cross section setting. Lines (concatenated collinear edges) from the triangulations are superimposed on the stimulus (the leftmost window). Also shown to the observer is either the middle or the right window. In this window we repeat the line, either horizontal ("method A"), or in the same orientation ("method B"). The task of the observer is to move the points orthogonal to the line (using a computer mouse as interaction device) such as to reproduce a cross–section of the pictorial relief (thus the direction of movement is the "depth" dimension). This is an adaptation of a method pioneered by Frisby et al.*[22]

All these methods allow one to check for surface integrity.[23] We have found no exceptions to the fact that coherent pictorial surfaces are produced by all methods at all times. This is perhaps remarkable since the sampling of surface locations is always done in random order, piecemeal fashion, thus the samples are expected to be independent. Apparently the observers have some data structure in their heads that is equivalent to an integral surface. Pictorial relief is very far from being anything like a "$2\frac{1}{2}$ D sketch".

2.2. Response Variation in Pictorial Relief

When we measure pictorial relief we find that the results are variable over subjects, tasks, viewing conditions and parametrically varied images (for instance photographs taken with various light source positions). Only rarely is pictorial relief quantitatively close to the fiducial scene.[24,25] (See figure 3.)

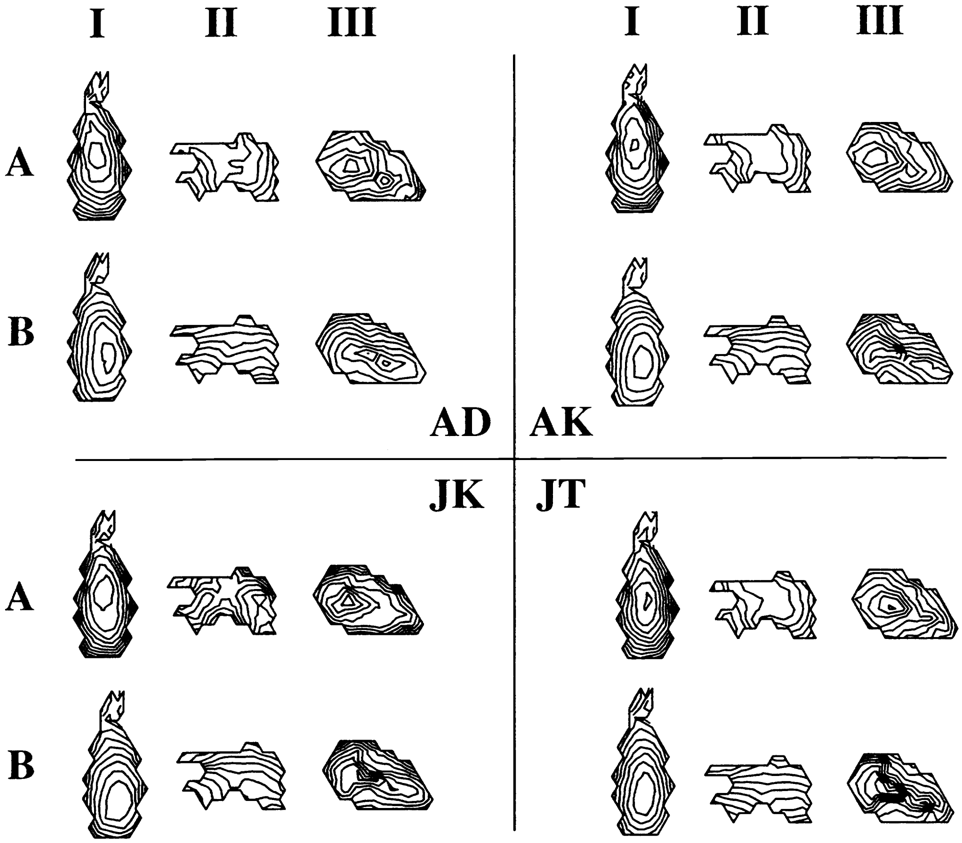

Figure 3. *Pictorial reliefs for four observers (AD,- AK, JK and JT), three stimuli (I, II and III) and two methods (A and B). The reliefs are indicated by way of equally spaced curves of equal depth. Notice the rather dramatic differences in picorial relief. Clearly not all these responses can simultaneously be "veridical" in the naive sense. Most likely none of them is, that would indeed be next to miraculous.*

When viewing conditions are varied the monocular cues are invariant. In such cases the pictorial reliefs for any given observer tend to be simply related. We find that the variation is almost fully explained through stretches of the depth domain then. A simple example is viewing with both eyes or monoculary. When one closes an eye while viewing a picture one actually witnesses the subsequent expansion of pictorial relief in the depth dimension. Such effects were already familiar to Leonardo.[26] They have been exploited by the optical industry.[27,28]

3. THE OBSERVER'S SHARE IN PICTORIAL RELIEF

3.1. Stratification of Perception

As indicated earlier, percepts (in this case pictorial reliefs) are due to two independent causes. One is the structure of the image in terms of the cues used by the observer (section 1.2). The other is the observer's share in resolving the ambiguity (section 1.3) left by the constraints implied by the cues. One might say that the former part is what is causally determined by the image (actually: the cues). This is a conceptually important *stratification of scene perception.* One wonders how it might be operationalized.

Since we typically don't possess knowledge of the fiducial scene we can't compare the pictorial relief (section 2.1) with the fiducial object directly. However, we can certainly compare two pictorial reliefs obtained from images due to the same fiducial scene, even if that fiducial scene must remain unknown. In many cases the images will even be identical and the different reliefs due to variation of psychophysical method, observer, and so forth. In this case those parts of the pictorial reliefs that are causally determined by the image should be equal, the two reliefs should be metameric or equal *modulo* an ambiguity transformation. This is something that can be tested, at least if one knowns the group of ambiguity transformations (section 1.3).

Problem is that the group of ambiguity transformations is not known. At this stage of scientific development we understand the groups of ambiguity transformations only for a few cases of simplified "shape from X" problems. In realistic cases of pictorial space we don't even know the cues actually employed by the observer. Thus we are in no position to delineate the ambiguity group from first principles.

A simple and very general line of reasoning to at least obtain *some* notion of what the nature of the ambiguity group might be like is the following. Notice that almost all cues as described in the literature let you *verify planarity of pictorial relief.* For instance, in traditional shape from shading the object is irradiated with a uniform collimated beam (the sun say). The surfaces are Lambertian so viewing direction is irrelevant. The only relevant parameter is the obliquety of the incident beam with respect to the local surface normal. The former is constant by assumption. Thus you obtain a uniform irradiance (featureless blob in the image) whenever the local surface normals are constant, i.e., when the surface is a planar one. Any deviation from planarity will induce shading and can thus be spotted. Similar reasoning applies to virtually all other cues. This is an important observation. It means that the ambiguity transformations should conserve planarity in pictorial space. They should also conserve the pencil of visual rays (not just the pencil as a whole, but in a ray–wise manner), because of the simple fact that the picture plane is (trivially!) conserved in a pointwise manner. This essentially constrains the ambiguity transforms to certain projective transformations though. In the case of orthographic projection (for several reasons this applies to pictorial perception in our experiments) the ambiguity transformations are limited to certain affinities, namely depth scalings and shears conserving the picture plane. With this reasoning we skip the need to solve dozens of complicated "shape from ..." problems. We may confidently wait for computer vision to catch up because we already have most of the answer in so far as it is relevant to the interpretation of psychophysical results.

The upshot of this is that a point $\{x, y, z\}$ (here $\{x, y\}$ denote Cartesian coordinates in the picture plane, whereas z denotes the depth) of pictorial relief in one case will correspond to a point $\{x, y, \alpha x + \beta y + \gamma z + z_0\}$ in a metameric case. The shift z_0 is irrelevant since pictorial relief is only known up to an arbitrary shift anyway. The parameters (α, β, γ) describe an arbitrary depth scaling combined with a shear. In order to check whether two pictorial reliefs are metameric we simply find the best values for $(\alpha, \beta, \gamma, z_0)$ in a least squares sense and check whether the residuals can be explained as noise. A multiple regression of the depths (for all image positions) of one case against the depth of the other case *and the image coordinates* performs exactly this. We call it "affine correlation". Comparison of pictorial reliefs by way of affine correlation automatically discounts any observer's share.

When two pictorial reliefs are found to have a high affine correlation (section 3.1) they are metameric and thus equal to the extent that they causally depend on the image structure (the cues). The remaining difference (depth scaling and shear) is exactly the "observer's share".

We have found many examples of variation over viewing conditions, psychophysical tasks and change of observer in which the resulting pictorial reliefs were *very different* indeed. The linear correlation between the depth values at corresponding points can be vanishing or even negative. In the great majority of cases we find that the affine correlation coefficients are very high (over 0.9) though. (See figures 4 and 5.) In such cases the differences are clearly due to the observer's share(s). Thus at least one pictorial relief (but very probably both) has to be quite unlike the fiducial scene. Thus the naive notion of "veridicality" (section 1.4) makes no sense.

3.2. Nature of the Observer's Share

What is the "observer's share" (section 1.5) like? Well, it is always an ambiguity transformation (section 1.3), thus a depth scaling combined with a shear. The nature of these two components is quite different though.

Depth scalings have been described as essential visual ambiguities by sculptors. The German sculptor Adolf Hildebrand[29] wrote a very articulate book on the subject in which he relates the resulting structure to æsthetic problems. He remarks that human observers easily confuse bas relief sculpture with sculpture in the round when the

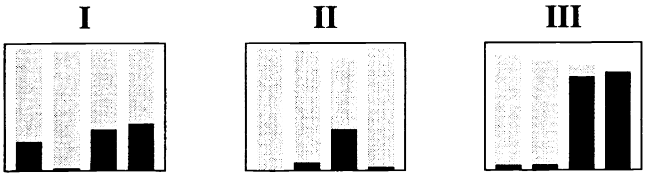

Figure 4. *Coefficients of determination for straight (black) and affine (gray) regression of depth values obtained with methods A and B. Results for observers AD, AK, JK and JT are plotted for stimuli I, II and III. The vertical extent of the boxes indicates the range (0. . . 1) of possible values for the coefficients of determination. In the majority of cases the coefficients of determination are insignificant for the straight comparison. They are invariably high for an affine regression though.*

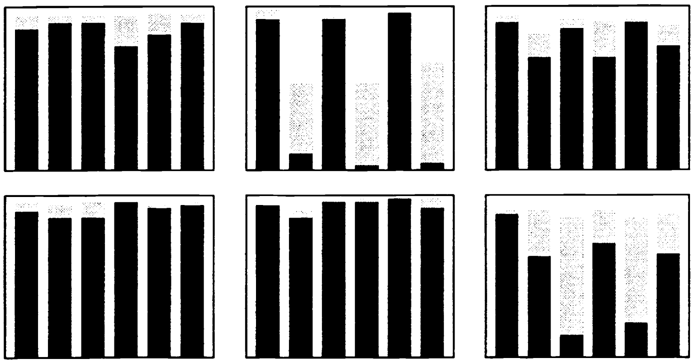

Figure 5. *Coefficients of determination for straight (black) and affine (gray) regression. Columns are for stimuli I, II and III, rows for methods A and B. Bars denote comparison of observers AD–AK, AD–JK, AD–JT, AK–JK, AK–JT, and JK–JT. Again, the vertical extent of the boxes indicates the range (0. . . 1) of possible values for the coefficients of determination. In several cases the coefficient of determination is very small for a straight comparison. Affine regression yields a high coefficient of determination though. In these cases the observer's shares must be very relevant indeed.*

circumstances are appropriate. Indeed, bas relief treatment has been common in modern western sculpture since the renaissance and is a very effective means of shape expression.

Shears involving depth but conserving the picture plane have not figured in the literature. They are interesting though because—different from the mere depth scalings—they allow for a change in surface orientation. (See figures 6 and 7.)

It is an intriguing fact that by selecting the appropriate shear *any local part of pictorial relief can be made frontoparallel.* In this sense the shears resemble rotations about frontoparallel axes. The difference is that a rotation will lead to changes of self–occlusion. For instance, I can change from an anterior to a posterior view of a statuette by rotating it about the vertical (assuming horizontal viewing direction). This is clearly out of the question in pictorial perception. The observer's share is unable to reveal parts of objects that are not in the image. The shears are indeed special in that they conserve self–occlusion yet allow change of surface orientation.

Physically those parts of objects that are frontoparallel have normal directions coinciding with the viewing direction. This relation changes in pictorial relief when shears are applied. As mentioned above *any* part of the relief can be made to appear frontoparallel. Notice that this can be interpreted as a (virtual) change of viewing direction, the virtual viewing direction being along the normal direction after application of the shear. This is indeed one possible interpretation of the observer's share: *Application of shears allows the observer to change the viewing direction of the "mental eye".* (See figure 8.) This is the nature of the oberver's selection from the metameric set of multiple visual worlds.

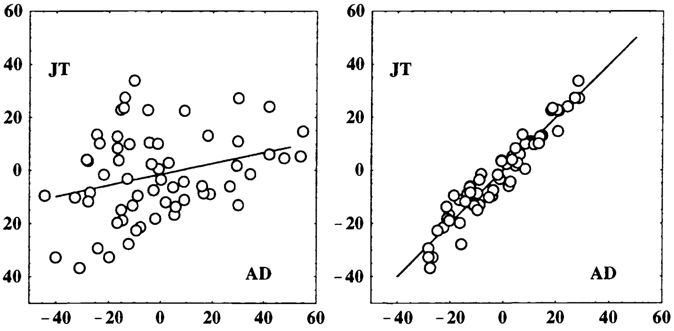

Figure 6. *Comparison of pictorial relief for observers AD–JT, stimulus III, task B, straight and affine regressions. The values along the axes are pictorial depth measured in units equivalent to pixels in the image plane. Notice that the origin of the axes is arbitrary. There is hardly any linear correlation, whereas the affine regression leads to a high coefficient of determination. This is a case where the observer's shares are spectacularly different. The pictorial reliefs are quite different but become comparable after correction by an appropriate ambiguity transformation.*

This seems to be exactly what happens in many of the cases studied by us. Often the pictorial objects have obvious "preferred views" (for instance when they are almost planar or have bilateral symmetry, etc.) and observers often tend to select shears that direct the mental eye to take such a canonical view, even if the photograph was taken from a generic direction.

4. CONCLUSIONS

4.1. Psychophysics of Scene Perception

There are many important and far reaching conclusions that apply to the psychophysics of scene perception. All conclusions somehow relate to the central topic of veridicality (section 1.4). Although the notion of veridicality is taken as "self evident" by the vision community at large, this is—we believe—so naive as to be useless.

The naive concept of veridicality is actually *unscientific* because it assumes that the observer possesses the uncanny extrasensory faculty to judge whether a percept is like the fiducial scene. The fiducial scene is only one of an infinite number of virtual scenes that all equally well explain image structure (in the sense of cues used by the observer). It is in no way singled out by the image structure. Thus observers would have to be able to read the mind of the author of the image in order to be able to come up with a veridical perception in the naive sense.

The naive concept of veridicality is also of *very limited use* because it can only be applied if knowledge of the fiducial scene is available. In many cases (almost all of the photographs you are confronted with in daily life) this is not the case. Human observers are not particularly bothered by this fact though. Illustrated newspapers and glossy magazines would go out of business if we didn't possess this remarkable tolerance. That human observers routinely entertain multiple visual worlds saves the day.

The naive concept of veridicality leads to *erroneous conclusions* on the basis of psychophysical results. For instance, when two observers yield different results it is conventionally concluded that at least one of them has to be the victim of illusionary perception. The reason is that it is assumed that only one perception can be non–illusionary, namely the one that is veridical (i.e., is like the fiducial scene). Likewise, if two psychophysical methods yield different pictorial reliefs, even for a single observer, it is conventionaly concluded that at least one method has to be flawed. Such examples can be multiplied. They lead to erroneous decisions (eg., when observers don't agree the study is worthless, when two methods yield different results one should figure out which one is the wrong one, etc.).

There must be many instances in the literature where apparently different results (vanishing correlation say) were actually identical in so far as causally dependent on the image structure. Thus one has to exert extreme care in the applications of even "clear–cut" conclusions drawn from the psychophysical literature.

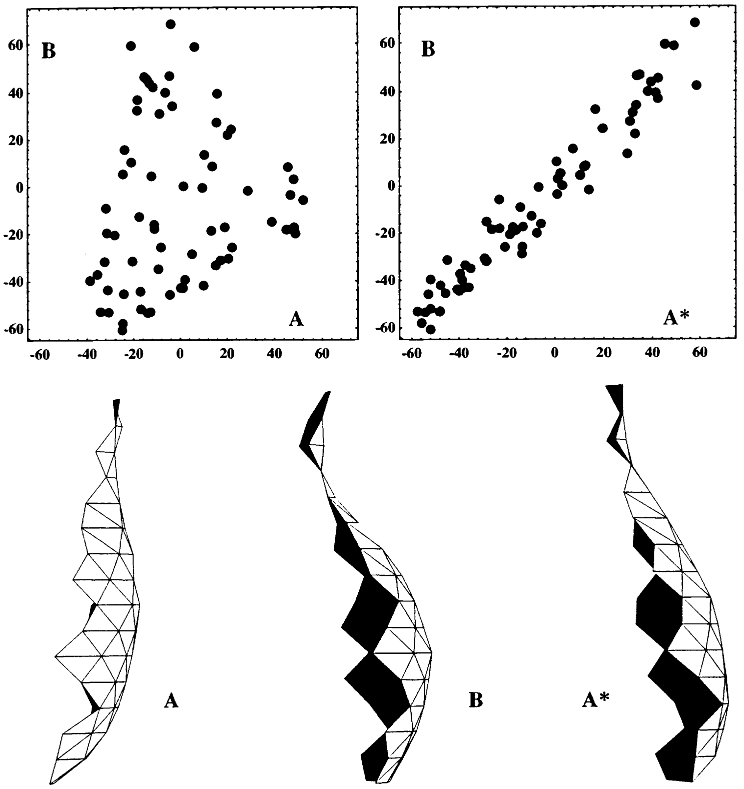

Figure 7. *Upper graphs: Straight linear regression of depths and affine regression for tasks A and B, observer AK, stimulus I. Again, the values along the axes are pictorial depth measured in units equivalent to pixels in the image plane. The origin of the axes is arbitrary. Lower figures: Effect of the ambiguity transformation. In this case the pictorial reliefs obtained for two (rather similar) methods for a single stimulus and a single observer turn out to be spectacularly different. The coefficient of determination is quite low unless we do an affine regression and thus discount the observer's share. In the bottom row it becomes evident that the observer's share consists in an appreciable tilting of the relief, that is to say, a redirection of the viewing direction of the "mental eye" of the observer.*

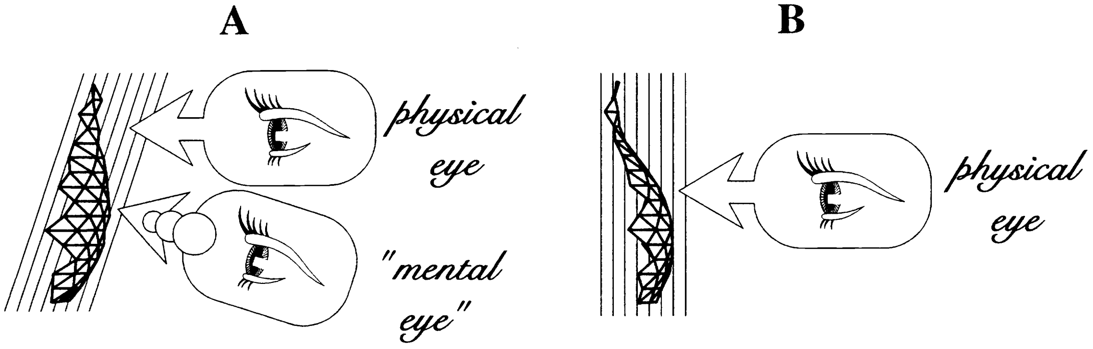

Figure 8. *The observer's share. This is the case illustrated in figure 7.*

4.2. Practical Implications

When pictures are used to instruct an agent one has to be aware that a picture does not specify the fiducial scene, but only the fiducial scene *modulo* the group of ambiguity transforms (section 1.3). Thus a picture specifies a large class of virtual scenes. Clearly, some of these virtual scenes are more likely than others. The observer will apply knowledge of ecological probability to arrive at a final judgment—or postpone judgment when that is possible. Such ecological knowledge is part of the observer's share. It is not specified by the image (section 1.4), thus the observer could well commit terrible mistakes from the perspective of an outsider (or scientist running the observer in a psychophysical experiment). Such mistakes are especially likely if the image is far from generic. The observer is likely to apply some genericity constraint and thus will be led astray. For instance, a broom stick photographed end–on is likely to be perceived as a ping pong ball, rather than as something extended in depth. In a frontal photograph of a face the observer is likely to perceive a nose of average length, even if the sitter happened to be Pinocchio, and so forth.

ACKNOWLEDGMENTS

This research was sponsored by NATO Scientific Exchange Grant CRG 92065.

REFERENCES

1. S. Geist, *Brancusi, the sculpture and drawings*, Harry N Abrams Inc., New York, 1975.
2. J. J. Koenderink, and A. J. van Doorn, "Geometrical modes as a general method to treat diffuse interreflections," *J.Opt.Soc.Am.* **73**, pp. 843–850, 1983.
3. K. Dana, B. van Ginniken, S. Nayar, and J. Koenderink, "Reflectance and Texture of Real–World Surfaces," IEEE Conf. on CVPR, p. 151, 1997.
   See also the internet database of Utrecht and Columbia University `http://www.cs.columbia.edu/CAVE/curet/` with dozens of BRDF's of natural materials.
4. G. Berkeley, *An essay towards a New Theory of Vision*, 1st printed 1709, in: *Philosophical works*, Introduction and notes by M. R. Ayers, J. M. Dent & Sons Ltd, London, 1975.
5. J. J. Gibson, *The Perception of the visual world*, Houghton Mifflin, Boston MA, 1950.
6. Euclid, *Optics*, original c. 300 BC, in: H. E. Burton, *The Optics of Euclid*, *J.Opt.Soc.Am.* **35**, pp. 357–72, 1945.
7. R. Riedl, *Biologie der Erkenntnis*, Deutscher Taschenbuch Verlag, München, 1988.
8. J. J. Koenderink, and A. J. van Doorn, "Illuminance critical points on generic smooth surfaces," *J.Opt.Soc.Am. A* **10**, pp. 844-854, 1993 .
9. J. J. Koenderink, A. J. van Doorn, Ch. Christou, and J. S. Lappin, "Perturbation study of shading in pictures," *Perception* **25**, pp. 1009–1026, 1996.
10. P. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The bas–relief ambiguity," *Proceedings 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Cat.No. 97CB36082), xviii+1117, pp. 1060–6, 1997.
11. J. J. Koenderink, and A. J. van Doorn, "Photometric invariants related to solid shape," *Optica Acta* **27**, pp. 981–996, 1980 .
12. J. J. Koenderink, and A. J. van Doorn, "The generic bilinear calibration-estimation problem," *International Journal of Computer Vision* **23**, pp. 217–234, 1997.
13. W. H. Ittelson, *The Ames Demonstrations in Perception*, Princeton University Press, Princeton NJ, 1952.
14. A. Byrne and D. R. Hilbert (eds), *Readings on Color. Vol. 1: The Philosophy of Color*, The M.I.T. Press, Cambridge MA, 1997.
15. J. D. Foley, A. van Dam, S. K. Feiner, and J. F.Hughes, *Computer Graphics, Principles and Practice*, 2nd ed, Addison–Wesley, Reading MA, 1990.
16. J. J. Koenderink, *Solid Shape*, The M.I.T. Press, Cambridge MA, 1990.
17. J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers, "Pictorial surface attitude and local depth comparisons," *Perception & Psychophysics* **58**, pp. 163–173, 1996.
18. J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers, "Surface perception in pictures," *Perception and Psychophysics* **52**, pp. 487–496, 1992.
19. J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers, "On so–called paradoxical monocular stereoscopy," *Perception* **23**, pp. 583-594, 1994.
20. J. J. Koenderink, and A. J. van Doorn, "Relief: Pictorial and otherwise," *Image and Vision Computing* **13**, pp. 321–334, 1995 .
21. J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers, "Depth relief," *Perception* **24**, pp. 115–126, 1995.

22. J. P. Frisby, D. Buckley, F. Bayliss, and J. Freeman J, "Integration of conflicting stereo and texture cues in quasi–natural viewing of a torso sculpture," *Perception* **24** Suppl., p. 136, 1995.
23. J. J. Koenderink, "Pictorial relief," *Phil.Trans.R.Soc.Lond. A* **356**, pp. 1071–1086, 1998.
24. J. J. Koenderink, A. J. van Doorn, Ch. Christou, and J. S. Lappin, "Shape constancy in pictorial relief," *Perception* **25**, pp. 155–164, 1996.
25. J. T. Todd, J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers, "Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces," *Journal of Experimental Psychology: Human Perception and Performance* **22**, pp. 695–706, 1996.
26. Leonardo da Vinci, in: J. P. Richter, and I. A. Richter, *The literary works of Leonardo da Vinci compiled and edited from the original MSS*, 2nd ed, 2 vols, Oxford University Press, Repr. Phaidon, London, 1970.
27. M. von Rohr, "Linsensystem zum einäugigen Betrachten einer in der Brennebene befindlichen Photographie," Kaiserliches Patentamt, Patentschrift Nr 151312, Klasse 42h, 1904.
28. Carl Zeiss Jena, "Instrument zum beidäugigen Betrachten von Gemälden u-dgl., das aus einer geraden Zahl gegen die Mittellinie des Objektraums um 45° geneigter Spiegel in oder außer Verbindung mit einem Fernrohrsystem besteht," Kaiserliches Patentamt, Patentschrift Nr 194480, Klasse 42h, Gruppe 34, 1907.
29. A. Hildebrand, *Das Problem der Form in der bildenden Kunst*, Strassburg, transl. M Meyer and R M Ogden 1945, *The problem of form in painting and sculpture*, G E Stechert, New York, 1893.