

Binocular disparity only comes into play when everything else fails;
a finding with broader implications than one might suppose.

Zygmunt Pizlo & Yunfeng Li

Department of Psychological Sciences, Purdue University, West Lafayette, IN

Robert M. Steinman

Department of Psychology, University of Maryland, College Park, MD

Abstract: This paper calls attention to research showing that binocular disparity, which is an effective cue to depth, plays a secondary role, at best, in the perception of 3D shape. This claim has implications both for how shape should be studied and how this unique perceptual property should be modeled. These issues are discussed from a historical perspective, which shows how the failure to appreciate the importance of the Gestalt grouping principle called “Figure-Ground Organization” led to many unfruitful efforts. It also calls attention to how this situation can be remedied.

Key Words: Binocular vision, 3D shape, simplicity principle, nativism/empiricism, figure-ground organization.

Corresponding author: Zygmunt Pizlo
pizlo@psych.purdue.edu

Introduction

Conventional, textbook wisdom in visual science says that binocular disparity, as one of the strongest, if not the strongest, depth cue, contributes significantly to the percept whenever the observer uses both eyes to look around, providing only that a disparity signal is detectable. Not everyone takes this textbook wisdom for granted, which is why this special issue was organized. In this paper, we discuss fundamental limitations of binocular disparity when it serves as a cue for shape, and derive from this discussion some critical comments about how shape, in particular, and visual perception, in general, should be studied at this point in time. We begin with a brief statement of our three main claims (hypotheses):

1. These days, it is not provocative to say that the perceptual representation of the shapes of 3D objects involves such geometrical properties of the objects as their symmetry, skeleton and volume (Marr, 1982; Biederman, 1987; Cornea et al., 2005; Palmer, 1999; Pizlo, 2008). Such geometrical properties, which give an object its shape, cannot be sensed or measured by either the haptic or visual system because these properties are “hidden”. One can touch the surface of an object, but not its skeleton, because the skeleton is inside the object. The same is true of volume because volume is behind the object’s visible surface. The symmetry of a 3D object can be assessed visually, but only if the object is transparent. Transparent objects are relatively rare in our environment and when an object is opaque, as most are, its back surfaces are hidden. So, more often than not, the visual information available on an observer’s retina is simply not sufficient to allow him to sense an object’s 3D symmetry directly. This means that binocular disparity cannot be used to measure the shape of a 3D object because it does not provide information about its 3D symmetry, skeleton and volume. This limitation also applies to other visual cues to depth such as shading, texture and motion parallax. One can touch the surfaces of objects, and one can use binocular disparity to perceive the orientation of their surfaces and their curvature, but perceiving these properties is not the same as perceiving the 3D shape of the object. So how do we manage to perceive the 3D shapes of objects, which we do so well? The 3D symmetry, volume and skeletons of 3D objects are *added* by the observer’s perceptual system by using *a priori* knowledge. These considerations make it possible to hypothesize that binocular disparity should have little or no effect, whatsoever, on the percept of an object’s 3D shape, despite the fact binocular disparity does affect the percept of a 3D object’s surfaces. Recent evidence supporting this hypothesis, and its implications, are described in this paper.
2. Most previous efforts in visual science have been directed to studying “effects” of binocular disparity and other depth cues on the observer’s percept under variety of stimulating conditions. We will describe how contemporary, as well as earlier, applications of this approach, which continues an empiristic tradition going back to John Locke (1690), has only limited theoretical significance despite the fact that many “effects” of binocular disparity, as well as other depth cues, have been demonstrated over the years. Correlations among a host of visual cues, including binocular disparity, and the percept of shape, color and motion, should be expected because many brain areas are known to be interconnected. So much so that there is probably an inexhaustible supply of new combinations of cues that could be studied. The significance of such past and future

research, however, can be questioned because the significant question in visual perception is not what percept or brain area is related to what cue or brain area, but, rather, how do percepts achieve veridicality (achieve “constancy”). Recall that a percept is said to be “veridical” when the percept of a given object’s property, such as shape, agrees with the object’s property “out there”. The percept is “constant”. It does not change when the viewing conditions change. The traditional explanation of the various perceptual constancies is empiristic. The veridicality of one’s percepts is based on learning. In the case of shape, one learns to take its surface orientation into account. This traditional empiristic approach is simply no longer good enough because it does not lead to models that make testable predictions about the most significant question in perception. Specifically, what mechanisms are responsible for the human’s capacity to perceive objects veridically. Note that this criticism of the goal of traditional research in visual science is not limited to the time-hallowed empiristic tradition in perception. It also applies to modern Bayesian models, in which *a priori* constraints, rather than the resulting percepts, are assumed to have been established empirically, i.e., through prior experience. The significant problem for the visual scientist is to discover the nature of the *a priori* knowledge that is used by the visual system to achieve veridical percepts. This kind of information has to be *discovered* in experiments and clearly specified in computational theories (models). It cannot simply be assumed as is commonplace, today. Note, however, that explaining veridical percepts by performing experiments designed to demonstrate that a percept is *not* affected by varying stimulating conditions (studying perceptual constancy) represents a paradigm shift. It is fundamentally different from the approach used by most prior and contemporary visual scientists. Some recent evidence, which illustrates the value of studying perceptual constancies, rather than “effects” of cue combinations, are also described in this paper.

3. The property called “shape” is unique. It has a superordinate position in visual perception. It provides a great deal of the reliable information we have about things “out there”. Shape is much more complex than other perceptual properties and its complexity allows shapes to be readily distinguished from each other. Furthermore, they can be distinguished regardless of the direction from which they are viewed. In other words, the human visual system achieves a high degree of “shape constancy” despite the fact, initially emphasized by Berkeley (1709), that the 2D information available on the retina is not sufficient to provide reliable perceptions of 3D objects in natural environments. Other perceptual properties can be important, too. Color can tell us, as it did our ancestors, that there is ripe fruit in a forest made up of green leaves. Motion, size and distance can also be important, but this information is much less complex than shape. Each can be described by at most 3 parameters. These properties, or cues, can give us useful information about where things are and what they are doing, but these “things” are the 3D objects “out there” known to us by their shapes derived from their 2D images on our retinas. Once this is appreciated, it becomes clear that binocular disparity, and all the other cues such as color, size, depth, direction and motion, serve to enhance the perception of objects that are given by another mechanism. These ancillary cues are subordinate to shape and they cannot be used to create the perceived shape of a 3D object whenever natural cues are present. This is the reason we claim in our title that the failure of binocular disparity to influence a percept, when the percept contains a 3D object, has

broader implications than one might suppose. This paper contains only a very brief explanation of this admittedly very provocative claim. It is developed in great detail in a book entitled, “*3D shape: its unique place in visual perception*” one of us recently published (Pizlo, 2008).

Binocular disparity has no role in shape perception: Recently, Pizlo, Li & Francis (2005) showed that the traditional view of the role of binocular disparity in 3D shape perception must be changed. According to the traditional view, perhaps represented best by Marr (1982), visual processing begins with solving the binocular correspondence problem and then computing a 3D representation of the visible surfaces of the 3D shape based on their binocular disparities. Binocular disparity is particularly useful because disparity, alone, without any other visual cues and before figure-ground organization, is able to produce the required 3D percept. From this perspective, binocular disparity is a low-level feature that can come into play before higher level features are established. This was demonstrated by Julesz (1960), when he used random dot stereograms to produce percepts of 3D shapes without providing any useful monocular information about the 3D shape that would be perceived. In everyday life, binocular disparity does not operate in isolation. Other cues like shading, texture and motion are often available, as well. In such realistic cases, the depth information provided by several different cues is combined (fused) to produce a more reliable and more veridical 3D percept (e.g., Landy et al., 1995). Furthermore, the visual system is also able to combine visual cues with *a priori* constraints, such as the smoothness and rigidity of surfaces (Ullman, 1984; Grimson, 1982).

Pizlo *et al.* (2005) described a completely different mechanism. They were encouraged to do this by the results of an experiment on stereopsis in which binocular disparity was put into conflict with simplicity constraints. They used line drawings of 3D polyhedral objects, rather than random dot stereograms. A stationary cube was projected to the left eye and a left-right oscillating cube was projected to the right eye. Binocular disparity, operating with these stimuli, should lead to the percept of a non-rigid 3D object, being stretched and compressed along the line of sight of the left eye, but a *a priori* simplicity constraints (symmetry and compactness), should lead to the percept of a rigid, left-right oscillating cube. The subject, after fusing the two images, perceived a rigid, the left-right oscillating cube, the percept expected if the *a priori* simplicity constraints determined the percept (see demo:

http://viper.psych.purdue.edu/pizlo_cubes/). This result shows that binocular disparity was completely ignored. Note, furthermore, that the subject did not experience any binocular rivalry. This result also shows that binocular disparity was ignored. These results can be explained either by assuming that (i) *a priori* simplicity constraints (such as symmetry of the perceived shape and its compactness) are applied to the left and right retinal images *independently* to produce two 3D interpretations that are averaged, or (ii) the left and right retinal images are averaged and a single 3D percept is produced by applying the simplicity constraints to this averaged image. The fact that we see 3D objects as 3D in the presence of diplopia (e.g., fixate at your finger while paying attention to a 3D object that is farther away), suggests that the first explanation is more plausible than the second.

Pizlo *et al.* (2005) continued this line of research by determining which modification of the cube stimulus would change the percept in such a way that it agreed with predictions derived from assuming that binocular disparity was operating. They did this by removing the three edges forming the Y junction in the cube, and disconnected one vertex of the cube from the remaining parts of the cube (see the demo). Once this was done, the isolated vertex was perceived as

moving in depth along the line emanating from the left eye, while the rest of the contour was perceived as oscillating left-right. The movement of the vertex was consistent with the operation of binocular disparity, but the movement of the contour was not. These results have two implications: (i) binocular disparity operates across different objects, but not within them, and (ii) within a given object, *a priori* simplicity constraints determine the 3D shape perceived and binocular disparity is not operative. In other words, binocular disparity can be used to establish spatial relations among different objects, when no effective *a priori* constraints are present, but binocular disparity does not contribute to the shape perceived when effective *a priori* constraints are present.

It follows that there are two quite different mechanisms for 3D space perception: in one, binocular disparity (and other depth cues, such as motion parallax) is central. In the other, binocular disparity is barely (if at all) used. These two mechanisms are incompatible in the sense that there does not seem to be a single cue combination mechanism for fusing data that can accommodate both. Once we assume, as most visual scientists do, that ours is the best of all possible visual systems, we can ask why this might be? Why should binocular disparity be excluded? One important reason for excluding binocular disparity from shape computations might be that even a very small amount of visual noise leads to very unstable reconstructions of 3D spatial relations (Chan et al., 1999).¹ Once a 3D reconstruction is unstable, it may be difficult, perhaps even impossible, to incorporate *a priori* constraints such as symmetry or compactness, into a 3D shape reconstruction. Shape reconstruction is likely to be much more reliable if the computations begin by applying the *a priori* constraints to the retinal image *immediately* after figure-ground organization is established (Pizlo, 2008). Binocular disparity (and other depth cues) may be ignored, altogether, whenever effective *a priori* constraints are available. Chan et al. (2006) have reported evidence that supports this claim.

Their subjects were tested in a shape constancy task with 9 classes of line-drawings of randomly-generated unfamiliar 3D objects. On each trial, the subject was presented with stereo-pairs of line-drawings of two stationary objects, one after another, and had to decide whether the shapes of these 3D objects seen when the pairs were fused were “same” or “different”. In a trial employing “same” shapes, the second object was the same as the first, except that it was rotated around the vertical axis by 90 deg. Recall that *shape constancy refers to the ability to see the shape of an object as the same despite changes in the viewing orientation of the object relative to the observer*. In a trial employing “different” shapes, the two objects, which were different, were generated randomly. There were 9 sessions, each session involving one class of objects. Examples of each of the 9 classes of objects are shown in Figure 1.

Some of these objects were symmetric (A,B,D), others were not. Some had planar contours (A,E,G and H), while others did not have *any* planar contour (B,C,F and I). In one of these stimuli (D), some contours were planar but other contours were not. Finally, some were perceived as having surfaces enclosing volume (A,B,D,E,F), whereas others either did not have surfaces, or the surfaces did not enclose any volume. Note that all 9 classes of objects had similar underlying 3D structure; they all were based on 16 vertices of a polyhedron (objects G, H and I, involved a subset of the 16 vertices).

¹ The only exception, where binocular disparity is robust in the presence of visual noise, is stereoacuity. But stereoacuity is spatially local. It only involves judgments about whether a target is in front or behind a reference, not how far in front or behind (McKee et al., 1990). Stereoacuity is not likely to be critical in 3D shape perception because shape is spatially global and involves judgments about ratios of distances.

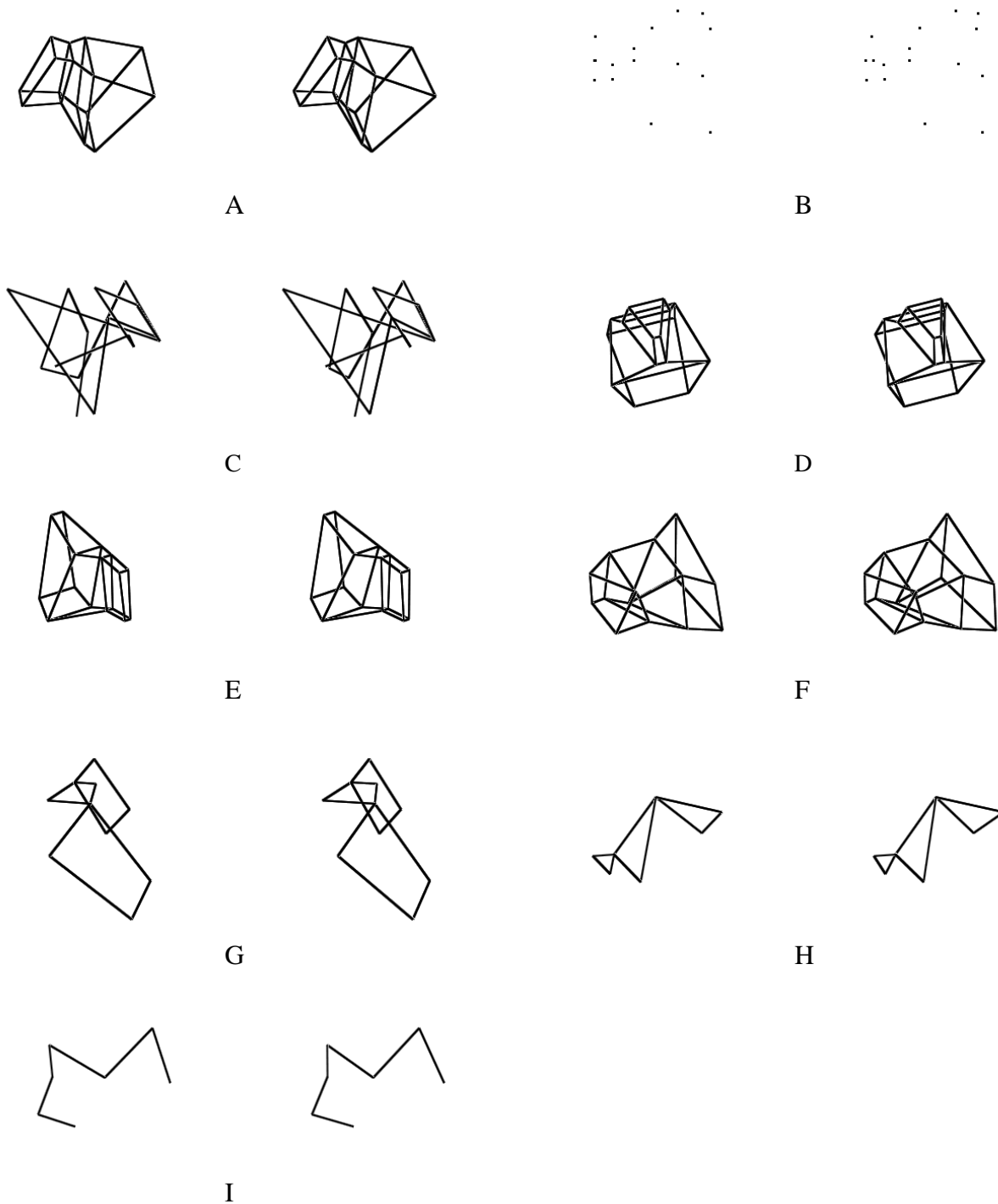


Figure 1. Stereoscopic images (for crossed fusion) of examples of stimuli in Chan et al.'s study (from Chan, et al., 2006). A – symmetric polyhedron; B – vertices of the polyhedron; C – polygonal line connecting the vertices of the polyhedron in a random order; D – symmetric polyhedron with some contours being non-planar; E – asymmetric polyhedron; F – asymmetric polyhedron with non-planar contours; G – three quadrilaterals, faces of the polyhedron; H – three triangles produced from the quadrilaterals in G; I – a polygonal line produced from the triangles in H. Shape constancy was most reliable with stimuli A, and least reliable with stimuli C, F and I.

All of the 9 classes of objects were perceived as three-dimensional, showing that the binocular disparity provided was an effective depth cue, but the degree of shape constancy achieved was quite different with the various classes of stimuli. The amount of shape constancy achieved varied with the number of shape properties (symmetry, planarity, volume) a given stimulus had. Shape constancy was most reliable with symmetric objects that had planar contours and volume (stimuli A) and shape constancy was not achieved when the stimuli were asymmetric, and had no planar contours or volume (stimuli C, F, H and I). It is important to note that shape constancy failed not only with polygonal line stimuli like C and I, which consisted of very simple features, i.e., line segments, but also with stimuli like F and H, which consisted of higher order features, i.e., contours. With polygonal line stimuli, poor performance might have been caused by the difficulty related to establishing the correspondence of individual lines segments across the two stimuli, rather than with establishing the 3D shape percept, itself. With several line segments in each stimulus, it may not be clear to the observer which line segment in one object “goes” with which in the other. Establishing correspondence will be much easier with objects like F and H, where contours of faces can be used. However, performance was close to chance level with all four types of stimuli C, I, F and H, which means that establishing correspondence of features was not the limiting factor. Shape constancy failed with these stimuli because binocular disparity was not sufficient to establish shape. Depth was perceived quite well with all of the stimuli, but shape was not. Clearly, *a priori simplicity* constraints are both *necessary and sufficient for achieving shape constancy, and binocular disparity is neither necessary nor sufficient for shape constancy*. The perception of depth provided by binocular disparity does not, and cannot, contribute to the veridical perception of shape.²

Chan et al., (2006); Pizlo, Li & Steinman, (2006); Li & Pizlo, (2007); and Pizlo (2008) used these psychophysical results to develop a computational model of shape constancy, in which *a priori* constraints they called symmetry, planarity of contours, minimum variance of angles, and maximal 3D compactness were used to reconstruct a 3D shape from a single 2D image of the 3D shape. The symmetry, planarity and minimum variance of angles constraints had been used before, but maximal compactness of the 3D shape was entirely new. It had never been used in a shape reconstruction model. Maximizing compactness is equivalent to maximizing the volume of an object, while keeping its surface area constant. It is also equivalent to minimizing surface area, while keeping the object’s volume constant. The 3D reconstructions produced by the model using shape constraints from a single 2D image were almost always veridical and they closely matched the percepts of human subjects when they were presented with the same stimuli. Depth cues, including binocular disparity were not used in any of the reconstructions. They are not needed. But one thing has to be made absolutely clear: the *a priori* constraints, listed above, can be applied *only after figure-ground organization has been established*. Specifically, the 3D reconstruction algorithm must be given the information about (i) which points in the image form edges, (ii) which edges and vertices in the image form contours of faces “out there”, (iii) which edges and vertices represent symmetric edges and vertices “out there”, and (iv) which edges and vertices define volume “out there”. It is necessary to provide all of this information because the symmetry, planarity, minimum variance of angles and compactness constraints are *shape* constraints. This means that they can only produce a percept of the 3D *shape* “out there” from the *shape* of the 2D image of the 3D object on the

² This experiment was also performed by using the kinetic depth effect (KDE), instead of binocular disparity to provide the perception of depth. The result with the KDE was the same, namely, it could not produce shape constancy when symmetry, planarity and volume were not provided in the stimuli (Pizlo & Stevenson, 1999).

retina. The Gestalt grouping principle, called “figure-ground organization” is required to produce the 2D shapes on the retina that can be used in these 3D reconstructions. Can binocular disparity contribute to establishing figure-ground organization? Probably. But the fact that we can easily see 3D shapes in photographs of shapes, suggest that the role of binocular disparity in figure-ground organization is secondary, at best.

The implications of the failure of binocular disparity to contribute to shape: Perhaps, the most significant undercurrent motivating our discussion of binocular disparity, depth cues and a priori constraints is what has been called the "nature-nurture problem", prominent in perception since Descartes (1637) and Locke (1690) took opposite sides in a 300 year long controversy that led to the Gestalt Revolution with Boring claiming in 1950 that Gestalt Psychology had died of success. The contributions of the Gestalt Psychologists to visual perception had influenced many visual scientists by 1950, which encouraged Boring to make this absurd claim, but a number of their most important insights were neither understood nor appreciated in their day and their importance has been lost during the ensuing years. Today, the tendency is to ignore the importance of the nature-nurture problem, simply saying that everyone acknowledges that nature and nurture always interact. This congenial, but superficial approach discourages serious analyses of the relative importance of learning and *a priori* constraints in visual perception. Many years have passed since the central arguments in the nature-nurture controversy have been discussed seriously, long enough for their nature and significance to have been forgotten in many quarters, particularly since Marr published his influential book in 1982. Marr’s approach to vision ignored one of the Gestalt Psychologists’ most important contributions, namely, the role of Figure-Ground Organization. Marr’s neglect shifted the emphasis in vision back from nativism to empiricism, a shift that has impeded progress in a number of ways for more than a quarter century. We believe that it is important to reconsider a number of the issues in this classical nature-nurture controversy, because further progress in visual science depends in large part on doing this now. In order to explain this claim, we will review some of the highlights in the development of visual perception since Gestalt “died of success”.

Gestalt psychology did not die of success as Boring claimed. Gestalt psychology actually died because it was ahead of its time. Its founders and most important protagonists (Wertheimer, Köhler and Koffka) did not live long enough to see many of the successes that would flow from their insights. The Gestalt Psychologists emphasized the role of simplicity constraints (their Law of Prägnanz, or simplicity principle) in determining the resulting percept, but “simplicity”, used intuitively as they did, was criticized as being too vague to have scientific utility. Attempts to define “simplicity” operationally only began formally after Shannon formulated Information Theory in 1948. Note that the operation of a simplicity principle in visual perception suggests that the visual system acts like a “control system” that solves “an optimization problem”. The formal theory of control systems, called “cybernetics”, was introduced when Wiener published his book in 1948. Unfortunately, neither Wertheimer nor Koffka lived till 1948. And even if they had, they probably did not have sufficient math or engineering background to take advantage of these developments even if they had recognized their significance. Köhler did have the necessary background, and lived long enough to see the advent of what has been called the “Cognitive Revolution”, but he was more interested in models rooted in physics. He devoted the last 20 years of his life trying to explain “figural aftereffects”, proposing that they were caused by brain currents established by “field forces”. He never expressed an interest in the new models rooted in electrical engineering.

The Cognitive Revolution was brought about by the introduction of Information Theory, Cybernetics and the construction of computers which made it possible to perform simulations. This revolution ultimately led to a resurgence of interest in Gestalt psychology by individuals called, “neo-Gestaltists”, but a group called, “neo-empiricists”, took the field over before the neo-Gestaltists made the connection between the simplicity principle and Information Theory. The neo-empiricists were led by Hebb who published his influential book in 1949. There were other counter forces during the same period, specifically a number of social psychologists led by Bruner and his colleagues who introduced the “New Look” in perception (Bruner, 1973). This group tried to demonstrate an important role of the subject’s motives, values, and learning history into low-level perceptual phenomena, including the perception of size and shape. A new emphasis on empiricism also entered the study of visual perception in this period. It was led by Ames and called, “Transactional Psychology” (Kilpatrick, 1961). It is famous for its “Distorted Room” which put the familiar size of common objects, such as people, in conflict with unnatural depth cues in the illusory environment in which they were viewed. The contribution of these neo-empiricist approaches to visual perception was rather modest, to say the least. All of these “theories” were abandoned within a decade or two following their introduction as the experiments purportedly supporting them were shown to be flawed.³

The neo-Gestalt approach to perception began shortly after 1948. It began when Hochberg & McAlister (1953) and Attneave (1954) published important papers attempting to operationalize “simplicity”. The fact that the mechanisms underlying visual perception are more likely to be innate than learned also received support from a series of experiments that started with Hochberg & Brooks (1962) who studied the ability of a young child to recognize 2D representations of 3D objects when they were seen for the first time. The revival of the Gestalt approach started with the study of shape, which should not be surprising once one recalls that the German word “Gestalt” means “form” (or shape). Although Max Wertheimer is credited as the founder of Gestalt psychology, the concept of the Gestalt itself was introduced by Christian von Ehrenfels in a famous paper entitled *Über Gestaltqualitäten* (“On Gestalt-qualities”) published in 1890. Christian von Ehrenfels also emphasized that the *Gestaltqualität* (form quality) was different from the elements making it up. Shape is not only different from its constituent elements, it is also more complex than all other visual properties, such as size, color and motion, taken together. Paradoxically, it is this complexity of shape that makes it relatively easy to perceive shapes veridically. From a computational point of view, however, there is no paradox because the complexity of shape makes it possible to apply very strong and effective simplicity constraints. These simplicity constraints, when they are applied to the 2D retinal shape produced by a 3D object “out there”, can lead to the veridical percept of the object’s 3D shape. But how is the shape in the 2D retinal image established?

The fact that establishing the 2D retinal shape that was produced by a 3D object involves specialized mechanisms was recognized quite early by the Gestalt psychologists (Wertheimer, 1923; Koffka, 1935). They called this specialized mechanism, “figure-ground organization”. The Gestalt psychologists also recognized that figure-ground organization could not be established without *a priori* constraints. These constraints were represented by a simplicity principle,

³ This observation is especially important now because we are currently experiencing a flurry of interest and activity in human vision within the machine vision community. Machine vision people always have had an empiristic bias. For some reason having a machine learn about the environment and then use this knowledge for a subsequent recognition seems more appropriate or natural to machine vision people than providing the machine with a general-purpose, intelligent program.

according to which *the percept was as simple as the stimulating conditions allowed*. Constraints are essential because a problem emerges when one wants to interpret the information present in the retinal image. The retinal image is inherently underconstrained, that is, there are always very many different 3D objects that could produce any given 2D retinal image. The visual system is faced with a considerable problem, namely, it has to be able to separate the shapes of objects from the backgrounds upon which these objects appear. The need for solving the figure-ground organization problem was recognized by the Gestalt psychologists very many years ago, but how it is actually accomplished by our visual system is yet to be explained (so much for Gestalt psychology “dying of success”). Note that the human visual system solves the figure-ground organization problem extremely well despite the fact that it has proven to be extremely difficult to understand how it manages to do this and to simulate its remarkable performance. By 1980, the machine vision community was seeking a way out of this impasse, having tried, unsuccessfully, to emulate human figure-ground organization for 25 years. A way to circumvent this impasse was provided by Julesz’ studies (1960, 1971) of stereopsis in which he introduced the use of random dot stereograms. Julesz’ demonstrations showed that percepts of 3D objects could be produced in the absence of any visible contours of these objects. Three-dimensional shapes could be perceived without forcing the visual system to solve the figure-ground organization problem. This development paved the way for David Marr (1982) to implement a major shift in visual science. It made it plausible for him to try to make progress in visual science without solving what had been its most pressing problem for many years, *viz.*, working out how the shapes of figures were organized and separated from their backgrounds. Note that Julesz’ demonstrations may have made it plausible to ignore the fundamental importance of the figure-ground organization but only if one ignored the fact that random dot stereograms have never been present in humankind’s ecologically-relevant environment. By 1980, Marr, who chose to ignore the importance of figure-ground organization in his treatment of human vision, had established himself, and his associates at MIT, as the leading group in visual science, and justly so. Marr, in his influential book published in 1982, emphasized that the goal of visual science is to understand and explain the perception of real objects in ecologically-valid environments, which included real images of real scenes. Marr also insisted on using computational models as the primary tool for evaluating our understanding of the underlying mechanisms. These two desiderata were universally accepted and remain the driving force for progress in visual science since Marr’s untimely death in 1980.

But, Marr’s willingness to ignore figure-ground organization was a huge mistake. He, like everyone else at that time, was impressed by Julesz’ demonstrations in which a 3D percept was produced without establishing figure-ground organization within each 2D retinal image. There were no monocular cues to the 3D shape, so there was no evidence of its 2D shape in the retinal image of either member of the pair of the random dot stereograms.⁴ The 3D percept was derived directly from binocular disparity. This encouraged Marr to claim that figure-ground organization is not needed, and went on to develop a theory that assumed that the human visual system does not make use of it (Marr, 1982).⁵ For Marr, the 3D shape percept was derived from

⁴ We do not claim that there are no monocular cues in random dot stereograms. After all, an observer can see the individual dots in each image. We claim, as Julesz did, that random dot stereograms do not provide monocular 3D cues.

⁵ This statement is often met with surprise. Marr did talk about establishing the primal and the final sketch in each of the retinal images, but Marr’s sketches had nothing, whatsoever, to do with figure-ground organization or with 2D shapes. Marr completely ignored figure-ground organization in his theory (1982, pp. 270-275). In his earlier work, Marr (1977) did analyze the relation between a class of 3D shapes, called generalized cones, and their 2D

local measurements of the orientations of visible surfaces of the objects. The surfaces, themselves, were computed from binocular disparity, supplemented by whatever other depth cues were available in a given scene.⁶ Binocular disparity and the other available depth cues were used by Marr as a substitute for *a priori* simplicity constraints that were deemed essential by the Gestalt Psychologists. Marr was an empiricist. In Marr, 3D shapes, including such properties of shape as symmetry, were learned, one after another, and placed in the observer's memory for future uses. The Gestalt Psychologists were nativists. According to the Gestalt Psychologists, 3D shape percepts are established spontaneously through the application of innate rules of perceptual organization, rules that make use of their simplicity principle. Marr was clearly on the right track when he proposed that a theory of a general purpose visual system, such as ours, will require formulating a computational theory of the underlying mechanisms, a theory that can lead to veridical percepts in the case of real images of real objects. His commitment to empiricism was both unnecessary and unfortunate. It diverted attention away from solving the figure-ground organization problem. The experiments putting 3D shape perception into conflict with binocular disparity, described above in this paper, call attention to the weaknesses inherent in the approach Marr adopted in 1982 when the figure-ground problem seemed to be intractable. They call attention to our continuing need to develop a theory of the *a priori* simplicity principle that is responsible for establishing figure-ground organization in the human visual system. We conclude by asking our readers to join us in this endeavor.

Acknowledgment: This research was supported by the National Science Foundation (grant # 0533968), US Department of Energy, and Purdue Research Foundation.

images for the purpose of reconstructing the skeletons of the 3D shapes from the 2D shapes present in the image. This analysis paved the way for Biederman's (1987) theory of recognition by components, but it was never pursued by Marr, himself.

⁶ Once the 3D visible surfaces are reconstructed from depth cues, they have to be grouped into individual objects, and separated from the background before they can be matched with 3D shape models preserved in the observer's memory. This grouping of surfaces into objects, however, is not what one means when referring to figure-ground organization. Figure-ground organization is performed on the basis of information present in the 2D retinal image; it consists of finding regions and contours and assigning contours to regions (Koffka, 1935). Once this is done, 2D shapes are established in the retinal image and they provide the basis for the percept of the 3D shapes of objects whose 2D shapes are given on the retinal surface. This is accomplished by using shape constraints (Pizlo, 2008). Reconstructing 3D surfaces from depth cues is not necessary for figure-ground organization, although it is necessary for shape reconstruction in Marr's (1982) theory. Conversely, 2D shape properties, such as symmetry of points and edges, closed contours, topological relations among contours and regions, are necessary for figure-ground organization, but not for shape reconstruction in Marr's (1982) theory. The 2.5D sketch in Marr's theory involves 3D local orientations of surfaces from the viewer's point of view. Discontinuities of surface orientation and of depth are present in the 2.5D sketch, but these discontinuities are not treated as 2D shapes in the image, but only as local features.

References

1. Attneave, F. (1954) Some informational aspects of visual perception. *Psychological Review*, **61**, 183-193.
2. Berkeley, G. (1709/1910) *A new theory of vision*. NY: Dutton.
3. Biederman, I. (1987) Recognition-by-components: a theory of human image understanding. *Psychological Review*, **94**, 115-147.
4. Boring, E.G. (1950) *A history of experimental psychology*. Englewood Cliffs, N.J. : Prentice Hall.
5. Bruner, J.S. (1973) *Beyond the information given: studies in the psychology of knowing*. NY: Norton.
6. Chan, M.W., Pizlo, Z. & Chelberg, D. (1999) Binocular shape reconstruction: psychological plausibility of the 8 point algorithm. *Computer Vision & Image Understanding*, **74**, 121-137.
7. Chan, M.W., Stevenson, A.K., Li, Y. & Pizlo, Z. (2006) Binocular shape constancy from novel views: the role of *a priori* constraints. *Perception & Psychophysics*, **68**, 1124-1139.
8. Cornea, N., Demirci, M. F., Silver, D., Shokoufandeh, A., Dickinson, S. & Kantor, P. (2005) 3D Object Retrieval using Many-to-many Matching of Curve Skeletons. *Proceedings, The International Conference on Shape Modeling and Applications*, MIT, June 2005, pp. 368-373.
9. Descartes, R. (1637/2001) *Discourse on method, optics, geometry, and meteorology*. (Translated by P.J. Olscamp) Indianapolis: Hackett.
10. Ehrenfels, C. von (1890) Über Gestaltqualitäten. *Vierteljahrschrift für Wissenschaftliche Philosophie*, **14**, 249-292.
11. Grimson, W.E.L. (1982) A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society, London*, **B298**, 395-427.
12. Hebb, D.O. (1949) *The organization of behavior*. NY: Wiley.
13. Hochberg, J. & Brooks, V. (1962) Pictorial recognition as an unlearned ability: a study of one child's performance. *American Journal of Psychology*, **75**, 624-628.
14. Hochberg, J. & McAlister, E. (1953) A quantitative approach to figural "goodness". *Journal of Experimental Psychology*, **46**, 361-364.
15. Julesz, B. (1960) Binocular depth perception of computer-generated patterns. *Bell System Technical Journal*, **39**, 1125-1162.
16. Julesz, B. (1971) *Foundations of cyclopean perception*. Chicago: University of Chicago Press.
17. Kilpatrick, F.P. (1961) *Explorations in transactional psychology*. NY: New York Univ. Press.
18. Koffka, K. (1935) *Principles of Gestalt Psychology*. New York: Harcourt, Brace.
19. Landy, M.S., Maloney, L.T., Johnston, E.B. & Young, M. (1995) Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, **35**, 389-412.
20. Li, Y. & Pizlo, Z. (2007) Reconstruction of shapes of 3D symmetric objects by using planarity and compactness constraints. *Proceedings of IS&T/SPIE Conference on Vision Geometry*, vol. **6499**.
21. Locke, J. (1690/1975) *An essay concerning human understanding*. Oxford: Clarendon.

22. Marr, D. (1977) Analysis of occluding contour. *Proceedings of the Royal Society of London*, **B197**, 441-475.
23. Marr, D. 1982) *Vision*. NY: W.H. Freeman.
24. McKee, S.P., Levi, D.M. & Bowne, S.F. (1990) The imprecision of stereopsis. *Vision Research*, **30**, 1763-1779.
25. Palmer, S.E. (1999) *Vision science*. Cambridge, MA: MIT Press.
26. Pizlo, Z. (2008) *3D shape: its unique place in visual perception*. Cambridge, MA: MIT Press.
27. Pizlo, Z. Li, Y. & Francis, G. (2005) A new look at binocular stereopsis. *Vision Research*, **45**, 2244-2255.
28. Pizlo, Z., Li, Y. & Steinman, R.M. (2006) A new paradigm for 3D shape perception. *Perception*, **35**, ECVF Abstract Supplement (p. 182).
29. Pizlo, Z. & Stevenson, A.K. (1999) Shape constancy from novel views. *Perception & Psychophysics*, **61**, 1299-1307.
30. Shannon, C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*, **27**, 623-656, 379-423.
31. Ullman, S. (1984) Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, **13**, 255-274.
32. Wertheimer, M. (1923/1958) Principles of perceptual organization. In: D.C.Beardslee & M.Wertheimer (Eds.) *Readings in Perception*, pp. 115-135. NY: D. van Nostrand.
33. Wiener, N. (1948) *Cybernetics*. Cambridge: MIT Press.