GESTALT ISOMORPHISM AND THE QUANTIFICATION OF SPATIAL PERCEPTION

Steven Lehar[1]

Introduction

Scientific theory is necessarily founded on certain philosophical assumptions. The philosophical underpinnings of science are not always apparent in mature sciences, where the correct philosophical groundwork has been established with such certainty that alternative philosophies appear too absurd for serious consideration. However in the case of sciences in an embryonic state of development, errors in the philosophical foundations can lead to grave errors in the science built on them. Nowhere is this more true today than in the science of mind and brain. Theories of visual perception can be separated into two classes, depending on their relation to a most significant philosophical distinction, i.e. the distinction between epistemological monism, or *naive realism*, versus epistemological dualism, or the *two-worlds hypothesis*. Therefore debates over the relative merits of opposing theories of vision are often at cross-purposes whenever the competing theories are founded on different philosophical assumptions. Such theories cannot be meaningfully compared without discussion of the differences in the underlying philosophy.

According to the naive realist view, the world we see around us is identfied as the objective external world, even though the limitations of our senses and the properties of light allow us to experience only a small subset of the properties of that world. In other words, the naive realist view holds that the world we see is the world itself. This is the natural intuitive understanding of vision that we accept from the earliest days of childhood. The problem with this view however becomes clear on consideration of the role of the eye as the sense organ of vision. For the flow of visual information occurs exclusively in one direction, from the world through the eye to the brain. If the brain is the organ of consciousness, then it cannot in principle experience the world directly, but only indirectly, in response to the two-dimensional images sent to it from the eyes. This fact is in conflict with our subjective experience of objects and surfaces outside of ourselves, because our conscious experience appears to escape the confines of our physical being, to extend into the external world beyond our sensory receptors. The causal chain of vision therefore refutes the naive realist view of vision, as explained by KÖHLER (1929). It is due to this naive realist view therefore that consciousness is often considered to be somehow mysterious,

forever beyond our capacity to comprehend, for there is no known physical mechanism that can possibly account for the external nature of visual experience.

The solution to this paradox was discovered centuries ago by Immanuel KANT (1781). KANT reasoned that we cannot actually experience the world itself as it is, but only an internal perceptual replica of the world. There are, in other words, two worlds of reality, the *nouminal* and the *phenomenal* world. The nouminal world is the objective external world, which is the source of the light that stimulates the retina. This is the world studied by science, and is populated by invisible entities such as atoms, electrons, and invisible forms of radiation. The phenomenal world is the internal perceptual world of conscious experience, which is a copy of the external world of objective reality constructed in our brain on the basis of the image received from the retina. The only way we can perceive the nouminal world is by its effects on the phenomenal world. Therefore the world we experience as external to our bodies is not actually the world itself, but only an internal virtual reality replica of that world generated by perceptual processes within our head.

Curiously this most central issue of vision has not received much attention in recent decades, and failure to understand this most significant issue has led to endless confusion in theories of visual representation. For the naive realist view suggests a very much simplified concept of the nature of the internal representation in vision. In the context of naive realism, introspective examination of the internal representation of vision, i.e. examination of the sensation within one's apparent head while viewing the world, reveals an abstract non-spatial entity as the internal code for external objects. This naive realist perspective therefore makes plausible many of the simplistic models of vision proposed over the centuries, and continues to cause confusion in modern neural network models of visual representation.

One reason for the persistent confusion on this issue is due to the fact that even the description of the causal chain of vision is somewhat ambiguous, since it can be interpreted in two alternative ways. Consider the statement that light from this page stimulates an image in your eye which in turn promotes the formation of a percept of the page. The ambiguity inherent in this statement can be revealed by the question "where is the percept?". There are two alternative correct answers to this question, although each is correct in a different spatial context. One answer is that the percept is up in your head, which is correct in the external or naive realist context of your perceived head being identified with your objective physical head, and since your visual cortex is contained within your head, that must also be the location of the patterns of energy corresponding to your percept of the page. The problem with this answer however is that no percept is experienced within your head where you imagine your visual cortex to be located. The other correct answer is that the percept of the page is right here in front of you where you experience the image of a page. This answer is correct in the internal spatial context of the entire perceived world around you being within your head. However the problem with this answer is that there is now no evidence of the objective external page that serves as the source of the light. The problem is that the vivid spatial structure you see before you is serving two mutually inconsistent roles, both as a mental icon representing the objective external

page which is the original source of the light, and as an icon of the final percept of the page; i.e. the page you see before you represents both ends of the causal chain. And our mental image of the problem switches effortlessly between the internal and external contexts to focus on each end of the causal chain in turn. It is this automatic switching of mental context that makes this issue so elusive, because it hinders a consideration of the problem as a whole.

I propose an alternative mental image to disambiguate the two spatial contexts. I propose that out beyond the farthest things you can perceive in all directions, i.e. above the dome of the sky, and below the solid earth under your feet, or beyond the walls and ceiling of the room you see around you, is located the inner surface of y-our true physical skull, beyond which is an unimaginably immense external world of which the world you see around you is merely a miniature internal replica. In other words, the head you have come to know as your own is not your true physical head, but only a miniature perceptual copy of your head in a perceptual copy of the world, all of which is contained within your real head in the external objective world. This mental image is more than just a metaphorical device, for the perceived and objective worlds are not spatially superimposed, as is often assumed, but the perceived world is completely contained within your head in the objective world (KOFFKA 1935, p. 27-36). The advantage of this mental image is that it provides two separate and distinct icons for the separate and distinct internal and external worlds, that can now coexist within the same mental image. This no longer allows the automatic switching between spatial contexts that tends to confuse the issue. Furthermore, this insight emphasizes the indisputable fact that every aspect of the solid spatial world that we perceive to surround us is in fact primarily a manifestation of activity within an internal representation, and only in secondary fashion is it also representative of more distant objects and events in the external world.

The Gestalt Principle of Isomorphism

Gestalt theory is founded on the philosophy of epistomological dualism (KÖHLER, 1938, pp 102-141,) For the illusory percepts studied by Gestalt theory, such as the moving light of the apparent motion effect, or the illusory surfaces of the Kanizsa and the Ehrenstein figures, are virtually indistinguishable from actual objects and surfaces in the visual world. These illusions therefore demonstrate that the brain is capable of constructing vivid spatial experiences that appear to consciousness as if they were raw sensations of real objects in the world. This in turn casts doubt on the objective reality of the non-illusory objects and surfaces in the visual world within which the illusory objects appear embedded, indicating that they too are internal copies of external objects and surfaces, rather than being those objects and surfaces themselves. It is this insight into the internal nature of the world we see around us that motivates the Gestalt principle of isomorphism. The theory of isomorphism was an outgrowth (KÖHLER, 1947, p. 57-60) of MÜLLER's psychophysical axiom (MÜLLER, 1896) which states that the subjective experience of perception cannot be of higher dimensionality than the neurophysiological state by which that experience is

encoded. More generally this concept is simply an expression of the materialist view that the properties of mind and consciousness are a direct consequence of electro-chemical interactions within the physical brain. Isomorphism differs subtly from MÜLLER's axiom in that it states explicitly what is only implied by MÜLLER, that in the case of structured experience, equal dimensionality between percept and representation implies similarity of structure or form (KÖHLER, 1947, p. 60-63). In the domain of color perception isomorphism is not controversial. Before the advent of neurophysiological confirmation, psychophysical experiments established the fact that the subjective experience of color can be reduced to the three dimensions of hue, intensity, and saturation. Perceived color therefore is of much lower dimensionality than the corresponding properties of physical light. It would be clearly absurd for example to propose that the neurophysiological mechanism underlying the experience of color should encode any less than three dimensions of information while producing three dimensions of color experience.

Curiously, in the realm of spatial perception this very obvious principle has not been accepted in contemporary psychology. Phenomenological examination of spatial perception reveals a world composed of solid volumes bounded by colored surfaces embedded in a spatial void. Every point on every visible surface is perceived at an explicit spatial location in three-dimensions, and all of the visible points on a perceived object like a cube or a sphere are perceived simultaneously in the form of continuous surfaces in depth. Furthermore, the perception of multiple transparent surfaces reveals that multiple depth values can be perceived at any spatial location. However proposed models of spatial perception very rarely allow for such an explicit representation of depth. MARR's 2 ½-D sketch (MARR, 1982) for example encodes the spatial percept as a two-dimensional map of surface orientations, like a two-dimensional array of needles pointing normal to the perceived surface. KOENDE-RINK & VAN DOORN (1976, 1980, 1982) propose a representation where each point in the two-dimensional map is labeled as either elliptic, hyperbolic, or parabolic, together with a number expressing the Gaussian curvature of the perceived surface at that point. TODD & REICHEL (1989) propose an ordinal map where each point in a two-dimensional map records the order relations of depth and/or orientation among neighboring surface regions. GROSSBERG (1987a, 1987b, McLOUGHLIN & GROSSBERG 1998) proposes a depth mapping based on disparity between two-dimensional left and right eye maps. None of these compressed representations are isomorphic with our subjective perception of a full volumetric depth world. In particular, all of these representations have a problem with encoding multiple surfaces at different depths, as in the perception of transparency, or encoding the volume of empty space that is perceived between the observer and a visible surface.

Naive Realism in Neural Network Theory

There are two possible approaches to the investigation of visual processing, a bottom-up approach by studying the elements of neurocomputation, and a top-down approach by studying the nature of the subjective experience of vision. Eventually these two approaches must meet somewhere in the middle, although to date, the gap between them remains as wide as ever. Neurophysiological studies of the visual cortex in experimental animals suggest a hierarchical visual representation composed of different levels of "feature detectors', i.e. cells that respond to the presence of particular features in the visual field. This concept of visual representation has served as a primary motivation behind many neural network models of vision (MARR, 1982; BIEDERMAN; 1987, HUBEL; 1988). Neural network theory suggests therefore that the internal visual representation is an abstraction or reduced dimensionality encoding of the objects and surfaces in the phenomenal world. The notion of perception by abstraction is supported by the practice of information compression, for example as used in digital image processing. The principle behind this kind of compression is the elimination of redundancy, either in the form of repeated values, or repeated sequences or patterns. For example images containing large regions of uniform brightness can be encoded in terms of the contrast along the edges bounding those regions, from which the brightness of the region can be reconstructed when necessary. In fact the representation of retinal ganglion cells appears to express exactly this kind of compressed image, since ganglion cells respond only along image edges, or spatial transitions of brightness in the visual field, and produce no response within regions of uniform brightness. ATTNEAVE (1954) suggests that the Gestalt principles of similarity, proximity, good continuation, symmetry etc. represent regularities in the visual world that offer an opportunity for information compression, to reduce to manageable proportions the overwhelming complexity of the visual world. For example a regular geometrical form can be encoded by its vertices only, which define the limits of the straight portions between them by the property of good continuation, just as the edges define the limits of the two-dimensional regions of uniform brightness that they separate. In some sense therefore the compressed representation encodes the same information as the full brightness image in which that information is expressed in redundant form, i.e. with complete boundaries separating regions explicitly painted in with repeated brightness values.

However the abstracted or reduced representation, while undoubtedly an essential component of perception, is not sufficient by itself to account for the nature of visual experience. For the subjective experience of perception is not of an edge image, but of a filled-in surface brightness image. If the retinal ganglion cells do in fact encode only transitions of brightness across image edges, then some process downstream of the retinal image must reverse the process and fill in the surface brightness values to account for the subjective experience of visual perception. In fact the identification of this constructive or generative aspect of perception represents one of the most significant contributions of Gestalt theory.

Perceptual Modeling v.s. Neural Modeling

One reason for the reluctance to accept a volumetric model of spatial perception is the apparent lack of neurophysiological evidence, given the two-dimensional structure of the visual cortex. KÖHLER himself felt it necessary to propose a radical model of neural representation in the form of an electric field theory (KÖHLER & HELD 1949) to account for the spatial nature of perception. According to field theory, the subjective percept of spatial structure is correlated with electric fields in the brain whose spatial pattern mirrors the spatial structure of the perceived world. KÖHLER's field theory was eventually disproven, at least in the specific formulation he proposed. Unfortunately the refutation of KÖHLER's field theory has been generally perceived as an indictment of the principle of isomorphism itself. However the validity of isomorphism stands independent of any specific neural hypothesis. If KÖHLER's field theory cannot be verified neurophysiologically, then some other mechanism of spatial representation must be sought that is isomorphic with the experience of spatial perception. If the neural network paradigm of visual representation in terms of spiking neurons and spatial receptive fields cannot be resolved with the principle of isomorphism, then it is our notions of neural representation that are in need of revision, not the principle of isomorphism. The question remains therefore how are we to model perception in the absence of a viable neurophysiological theory to supply the basic elements or building blocks for a model of perception?

I propose a *perceptual modeling* approach, i.e. to model the percept as observed subjectively rather than the neurophysiological mechanism by which it is supposedly subserved. In other words the perceptual model should be expressed in terms of solid volumes bounded by colored surfaces embedded in a spatial void, as observed in visual experience. This perceptual modeling approach must eventually converge with theories of neural representation, at which point it will be possible to relate the perceptual variables of color and shape to neurophysiological variables such as voltages or spiking frequencies as required. In fact, until a mapping is established between subjective experience and the neurophysiological state, a perceptual model is the only valid model to match to psychophysical data, which explicitly measures the subjective experience of perception rather than the corresponding neurophysiological state.

A Quantitative Phenomenology

Given the insights developed above, the dimensions of conscious experience can be established by direct phenomenological observation, just as were the dimensions of color perception. Since colored surfaces can be perceived at any location through a range of depths, and since transparent surfaces can be perceived simultaneously at multiple depths, the data structure required to encode the information of spatial perception must involve a volumetric manifold representing external space. Every point or region in that manifold can be in one of two states, transparent or opaque, and regions that are in the opaque state also take on a three-dimensional color value expressed in terms of hue, intensity, and saturation. The presence in this manifold of an opaque region encoding a particular color value is therefore by definition equiva-

lent to a subjective experience of a colored surface at the corresponding location in phenomenal space, whether that experience is perceptual, i.e. a veridical effigy of an external surface, or illusory as in the case of dreams or hallucinations. This is exactly the model of spatial perception suggested by KANT when he says "*On the occurence of a color-sensation* [one's mind] *reacts by producing a perceptual experience in which one is immediately presented with a color as pervading a certain region at a certain external position. All the regions which a color can ever be presented to one as occupying ... constitute a single three-dimensional spatial system.*" (BROAD, 1978, p. 29).

Given this kind of explicit spatial representation of subjective experience, the function of visual perception can now be expressed as a transformation from the two-dimensional visual input (or pair of two-dimensional images in the binocular case) to a solid three-dimensional volumetric representation of the spatial percept generated by that input. Whatever the neurophysiological reality of the perceptual mechanism, at least this information must be encoded neurophysiologically to account for the subjective experience of spatial perception. Merely expressing the problem in these terms eliminates a number of commonly accepted models of spatial representation.

Boundedness

This kind of phenomenological analysis of spatial perception immediately raises several fundamental issues about the required representation. One issue is the question of boundedness, i.e. how an explicit spatial representation can encode the infinity of external space in a finite volumetric system. The solution to this problem can be found by inspection. For phenomenological observation reveals that perceived space is not infinite, but is bounded. This can be seen most clearly in the night sky, where the distant stars produce a dome-like percept that presents the stars at equal distance from the observer, and that distance is perceived to be less than infinite. The lower half of perceptual space is usually filled with a percept of the ground underfoot, but it too becomes hemispherical when viewed from far enough above the surface, for example from an airplane or a hot air balloon. The dome of the sky above, and the bowl of the earth below therefore define a finite approximately spherical space (HEELAN 1983) that encodes distances out to infinity within a representational structure that is both finite and bounded. While the properties of perceived space are approximately Euclidean near the body, there are peculiar global distortions evident in perceived space that provide clear evidence of the phenomenal world being an internal rather than external entity.

Consider the phenomenon of perspective, for example how railroad tracks viewed in perspective appear to converge to a point in the distance. The reason why they converge has nothing to do with their objective geometrical arrangement, for parallel lines neither converge, nor do they meet at a point. However in perceived space the tracks are observed both to converge and to meet at a point, and that point is perceived at a finite distance beyond which the tracks are no longer represented. This

property of perceived space is so familiar in everyday experience as to seem totally unremarkable. And yet this most prominent violation of Euclidean geometry offers clear evidence for the non-Euclidean nature of perceived space. For the two rails are perceived to be straight and parallel throughout their length, even though they are also perceived to meet at a point up ahead and behind, while at the same time passing to either side of a percipient standing between them. The tracks must therefore in some sense be perceived as being bowed, and yet while bowed, they are also perceived as being straight. This can only mean that the space itself must be curved.

The curved properties of perceived space have been quantified in psychophysical experiments dating to observations by HELMHOLTZ (1925). Subjects in a dark room were presented with a horizontal line of point lights at eye level in the fronto-parallel plane, and instructed to adjust their displacement in depth until they were perceived to lie in a straight line in depth. The resultant line of lights curves inwards towards the observer, the amount of curvature being a function of the distance of the line of lights from the observer. The HILLEBRAND-BLUMENFELD alley experiments (HILLEBRAND 1902, BLUMENFELD 1913) extended this work with different configurations of lights, and mathematical analysis of the results (LUNEBURG 1950, BLANK 1958) characterized the nature of perceived space as Riemannian with constant Gaussian curvature (see GRAHAM 1965 and FOLEY 1978 for a review). In other words, perceived space bows outward around the observer, as seen in the bowed railway tracks.
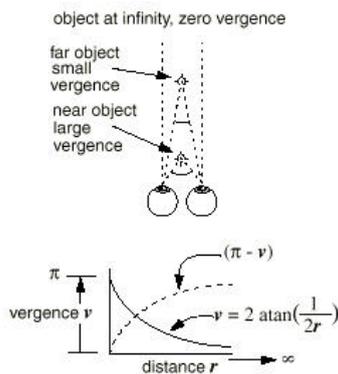
A   A vergence representation maps Euclidean distance into a finite bounded range.
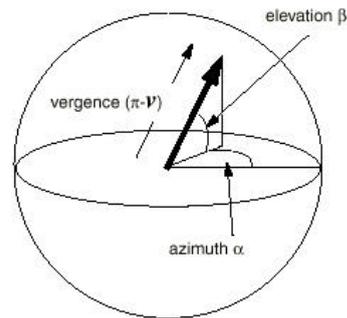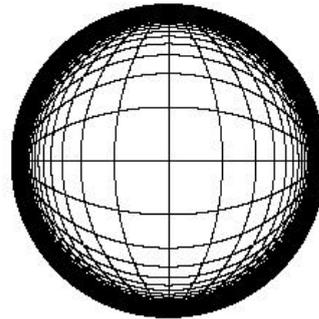
B   In a polar coordinate system the vergence measure of radial distance maps the infinity of Euclidean space into a bounded spherical representation. The outer surface of the sphere represents perceptual infinity.
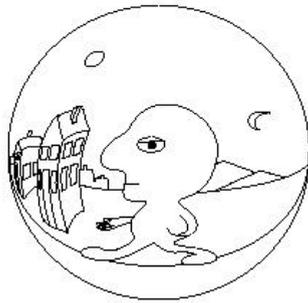
C

The perceptual representation of a man walking down a road.

D

The perceptual reference grid representing parallel lines at equal vertical and horizontal intervals.

Figure 1: a-d

The observed warping of perceived space is exactly the property that allows the finite representational space to encode an infinite external space. This property is achieved by using a variable representational scale, i.e. the ratio of the physical distance in the manifold relative to the distance in external space that it represents. This scale is observed to vary as a function of distance from the center of the manifold, such that objects close to the body are encoded at a larger representational scale than objects in the distance, and beyond a certain limiting distance the representational scale, at least in the depth dimension, falls to zero, i.e. objects beyond a certain distance lose all perceptual depth. This is seen for example where the sun and moon and distant mountains appear as if cut out of paper and pasted against the dome of the sky.

LEHAR & McLOUGHLIN (1998) propose a transformation to perceptual space using a polar coordinate system centered on the percipient, in which azimuth and elevation angles are preserved, but the radial distance is encoded in terms of vergence, or angle of convergence between eyes in a binocular system. In other words, point $P(\alpha,\beta,r)$ in Euclidean space is transformed to point $Q(\alpha,\beta,(\pi-v))$ in perceptual space, where α and β represent azimuth and elevation angles, while the radial distance r is compressed to the vergence representation v by the equation

$$v = 2\ atan(1/2r)$$

The vergence measure maps the infinity of Euclidean distance to a finite bounded range, as suggested in Figure 1a. Since azimuth and elevation angles are also closed dimensions, this transformation maps the infinity of Euclidean space into a finite

spherical space as suggested in Figure 1*b*. Figure 1*c* shows how such a compression of the depth dimension would encode the visual space around a man walking down a road. The fact that the distortion of this space is not immediately apparent to the percipient is explained by the fact that the percipient's sense of scale is itself distorted along with the space. For example the vertical and horizontal grid lines depicted in Figure 1*d* would be perceived to be straight and parallel, and separated by uniform intervals. If the reference grid of Figure 1*d* is used to measure lines and distances in Figure 1*c*, the bowed line of the road on which the man is walking is aligned with the bowed reference grid, and therefore is perceived to be straight. Likewise, the vertical walls of the houses in Figure 1*c* bow outwards away from the observer, but in doing so, they follow the curvature of the reference grid in Figure 1*d*, and are therefore perceived to be both straight and vertical. Similarly, the houses in Figure 1*c* would be perceived to be of approximately the same size and depth, although the farther houses are experienced at a lower perceptual resolution. This distortion of the perceptual reference scale accounts for the paradoxical but familiar property of perceived space, whereby more distant objects are perceived to be both smaller, and yet at the same time to be undiminished in size. This corresponds to the difference in subjects' reports, depending on whether they are given *objective* v.s. *projective* instructions (COREN, WARD, & ENNS, 1979. p. 500) in how to report their observations, showing that both types of information are available perceptually.

This "picture-in-the-head" or "Cartesian theatre" concept of visual representation has been criticized on the grounds that there would have to be a miniature observer to view this miniature internal scene, resulting in an infinite regress of observers within observers. PINKER (1984, p. 38) points out however that there is no need for an internal observer of the scene, since the internal representation is simply a data structure like any other data in a computer, except that this data is expressed in spatial form. The little man at the center of this spherical world therefore is not a miniature observer of the internal scene, but is itself a spatial percept, constructed of the same perceptual material as the rest of the spatial scene, for that scene would be incomplete without a replica of the percipient's own body in his perceived world.

Brain Anchoring

Another issue that must be addressed involves the subjective impression that the phenomenal world appears to rotate relative to your perceived head as your head turns relative to the world. This suggests that the internal representation of external objects and surfaces is not anchored to the tissue of the brain, as suggested by current concepts of neural representation, but is free to rotate coherently relative to the neural substrate, as suggested in KÖHLER's field theory. This issue of brain anchoring is so troublesome that it is often cited as a counter-argument for an isomorphic representation, since it is difficult to conceive of the solid spatial percept of the sur-

rounding world having to be reconstructed anew in all its rich spatial detail with eve-ry turn of the head (GIBSON, 1966, O'REGAN, 1992).

However an argument can be made for the adaptive value of a neural representa-tion of the external world that could break free of the tissue of the sensory or cortical surface in order to lock on to the more meaningful coordinates of the external world, if only a plausible mechanism could be conceived to achieve this useful property. The issue therefore is whether we have enough knowledge about the theory of in-formation processing systems to make a judgement about the plausibility of such a rotation invariant representation of spatial structure. The history of psychology is replete with examples of plausibility arguments based on the limited technology of the time which were later invalidated by the emergence of new technologies. The outstanding achievements of modern technology, especially in the field of informa-tion processing systems, might seem to justify our confidence to judge the plausibi-lity of proposed processing algorithms. And yet, despite the remarkable capabilities of modern computers, there remain certain classes of problems that appear to be fundamentally beyond the capacity of the digital computer. In fact the very problems that are most difficult for computers to address, such as extraction of spatial structu-re from a visual scene especially in the presence of attached shadows, cast sha-dows, specular reflections, occlusions, perspective distortions, as well as the prob-lems of navigation in a natural environment, etc. are problems that are routinely handled by biological vision systems, even those of simpler animals. On the other hand, the kinds of problems that are easily solved by computers, such as perfect re-call of vast quantities of meaningless data, perfect memory over indefinite periods, detection of the tiniest variation in otherwise identical data, exact repeatability of e-ven the most complex computations, are the kinds of problems that are inordinately difficult for biological intelligence, even that of the most complex of animals. It is the-refore safe to assume that the computational principles of biological vision are fun-damentally different from those of digital computation, and therefore plausibility ar-guments predicated on contemporary concepts of what is computable are not appli-cable to biological vision.

Indeed many of the most difficult aspects of vision are exactly those that were characterized by the Gestalt movement. A central focus of Gestalt theory was the is-sue of invariance, i.e. how an object, like a square or a triangle, can be recognized re-gardless of its rotation, translation, or scale, or whatever its contrast polarity against the background, or whether it is depicted solid or in outline form, or whether it is de-fined in terms of texture, motion, or binocular disparity. The ease with which these invariances are handled in biological vision suggests that invariance is fundamental to the visual representation. Even in the absence of a neural model with the required properties, the invariance property can be encoded in a perceptual model. In the ca-se of rotation invariance, this property can be quantified by proposing that the spa-tial structure of a perceived object and its orientation are encoded as separable vari-ables. This would allow the structural representation to be updated progressively from successive views of an object that is rotating through a range of orientations. However the rotation invariance property does not mean that the encoded form has

no defined orientation, but rather that the perceived form is presented to consciousness at the orientation and rate of rotation that the external object is currently perceived to possess. In other words, when viewing a rotating object, like a person doing a cartwheel, or a skater spinning about their vertical axis, every part of that visual stimulus is used to update the corresponding part of the internal percept even as that percept rotates within the perceptual manifold to remain in synchrony with the rotation of the external object. The perceptual model need not explain how this invariance is achieved computationally, it must merely reflect the invariance property manifest in the subjective experience of perception. The property of translation invariance can be similarly quantified in the representation by proposing that the structural representation can be updated from a stimulus that is translating across the sensory surface, to update a perceptual effigy that translates with respect to the representational manifold. This accounts for the structural constancy of the perceived world as it scrolls past a percipient walking through a scene, with each element of that scene following the proper curved perspective lines as depicted in figure 1*d*, expanding outwards from a point up ahead, and collapsing back to a point behind, as would be seen in a cartoon movie rendition of figure 1*c*. Whatever the computational mechanism behind this remarkable performance, these are the observed properties of the spatial percept.

The fundamental invariance of such a representation offers an explanation for another property of visual perception, i.e. the way that the individual impressions left by each visual saccade are observed to appear phenomenally at the appropriate location within the global framework of visual space depending on the direction of gaze. This property can be quantified in the perceptual model by proposing that the sensory image from the retina is copied onto the front surface of the eye of the perceptual homunculus, from whence that image is projected outward into perceived space in the direction of gaze, taking into account eye, head, and body orientation relative to the perceived world. Proprioceptive and kinesthetic information are used to update the body posture and orientation of the perceptual effigy of the body including the ocular orientation, to ensure that the retinal projection occurs in the appropriate direction in perceived space. In the case of binocular viewing, the projections from the two eyes are crossed in perceptual space, where their intersection in depth defines the three-dimensional binocular percept, as suggested by the projection field theory of binocular vision (BORING, 1933; CHARNWOOD, 1951; KAUFMAN, 1974; JULESZ, 1971; MARR & POGGIO, 1976).

The percept of the surrounding environment therefore serves as a kind of three-dimensional frame buffer expressed in global coordinates, that accumulates the information gathered in successive visual saccades and maintains an image of that external environment in the proper orientation relative to a spatial model of the body, compensating for body rotations or translations through the world. Portions of the environment that have not been updated recently gradually fade from perceptual memory, which is why it is easy to bump one's head after bending for some time under an overhanging shelf, or why it is possible to advance only a few steps safely after closing one's eyes while walking. Given the rotation invariance of the represen-

tation described above, it is immaterial whether the body percept rotates relative to a static world percept as suggested above, or whether the body or head percept remains fixed as the world percept rotates around it, either way would be isomorphic to the subjective experience.

The neurophysiological studies of the cortex using single cell recordings might appear to be inconsistent with the non-anchored representation proposed here. However the only cortical areas which are clearly defined spatial maps are the primary areas, such as the primary visual and somatosensory cortices. Cells in the higher cortical areas, while still somewhat topographic, exhibit progressively reduced spatial specificity, and in the highest level "association cortex" areas cells appear to lose all detectable spatial organization. This is exactly the property that would be expected in a non-anchored representation that is coupled in hierarchical stages to a brain-anchored map. Indeed the location of the parietal cortex between visual and somatosensory areas would suggest its function should be to associate the sensory-surface-mapped areas of vision and touch. But the spaces defined by the surface of the skin and the visual image on the retina can only be meaningfully related in a fully spatial context and by way of a non-anchored representation. It should come as no surprise that non-anchored patterns of activation in the cortex have not been detected in single-cell recordings, since the very nature of the brain-anchored electrode is predicated on an assumption of a brain-anchored representation.

Amodal Perception

There is another aspect of perception whose significance was recognized by Gestalt theory, but receives little mention in the contemporary literature. This is the phenomenon of amodal perception, or the perception of spatial structure that is not associated with any particular sensory modality. For example a book lying on a table is perceived to lie on a complete table top whose surface is continuous under the book, even though there is no sensory stimulus corresponding to the occluded portion of that surface. The hidden rear faces of objects are also perceived amodally, as observed by GIBSON (REED, 1988) and the Gestaltists (KANIZSA, 1979; ARNHEIM, 1969, p. 86). For example a sphere is not perceived as the hemisphere presented by its visible surface, but is experienced as a complete sphere, even though the percipient is also aware that the rear surface is hidden from view. Similarly, an object partially occluded by a foreground object is perceived to be complete behind the occluder. These phenomena indicate that it is possible to perceive spatial structure in the absence of physical stimulation, although the resulting percept exhibits a curious invisible character. Nevertheless, the spatial reality of such amodal percepts can be easily demonstrated by the ease with which a person can reach behind a sphere or cylinder and indicate with their palm the exact location and surface orientation of different parts of the hidden rear surface based exclusively on the view of the visible front surface. In order to account for this property another state must be defined in the perceptual manifold to represent volumes of solid matter in the absence of explicit visual stimulation. A percept of a sphere would therefore be represented

as a visible hemispherical front face, and this percept in turn would stimulate the activation of an invisible spherical volume in the perceptual manifold corresponding to the amodal percept of the whole sphere. This spatial completion mechanism can be formulated on the assumption that the visible portion is taken as a representative sample of the object as a whole, and therefore in the absence of contradictory evidence, the rear face is completed to match the front, i.e. performing a completion by symmetry. The volumetric spatial representation offers a computational framework that facilitates the detection of symmetry because a symmetry detection mechanism located at the center of curvature of the modal surface percept would be in a unique position to recognize, and therefore to complete the symmetry of the spherical form. This idea generalizes the concept of closure to include closure in depth, or a tendency to perceive objects as complete solid forms, a notion that lies at the very heart of Gestalt theory, from which the theory derives its name. A cylindrical object like a pillar would be represented as a hemi-cylindrical front surface expressed in modal terms, and that percept in turn would complete by symmetry to produce an invisible cylindrical core to match the curvature of the front surface. Any portion of this pillar that is occluded by a foreground object would thereby lose a portion of its modal front surface in the perceptual space, but the amodal cylindrical percept would complete across the occlusion by the principle of good continuation. The amodal structure therefore represents the object as a whole in a format that is independent of any particular sensory modality. This allows a variety of sensory stimuli to contribute to a single spatial percept, as was demonstrated by GALLI (1932) who showed that a stroboscopic motion stimulus composed of different sense modalities, e.g. light and sound, or light and contact, are perceived as a single moving object.

Perception Outside the Visual Field

The model developed above suggests that perception of visual space includes a percept of the world outside of the visual field, including the world behind the head. In other words, the head is treated as an occluder of the world behind the head, and the final percept is of a spherical space surrounding the body, only part of which corresponds to the visual field. Parts of the visual world that are currently outside of the visual field are experienced amodally, i.e. in the absence of a vivid impression of color and visual detail. However the world behind the head is experienced as a spatial structure, as can be demonstrated with a backwards step. A step (whether forwards or backwards) requires an accurate knowledge of the height and orientation of the ground at the point of contact. This becomes evident whenever a step encounters an unexpected change in surface height or orientation, even of as little as an inch or two, which inevitably results in a stumble. A backwards step without a stumble therefore indicates that the stepper has knowledge of these parameters within about an inch or two. The present model suggests that surfaces in the scene are extrapolated from their visible portions in the visual field into the unseen portion of the perceptual field in much the same manner as the amodal completion of the hidden rear faces of objects. For example the walls and ceilings of a hallway would

be completed perceptually behind the observer, as would such regular features as a handrail. This would explain how it is possible to accurately grab a handrail, pole, or surface at a point well outside of the visual field while viewing only the visible portion of the object. Both GIBSON (REED 1988) and the Gestaltists (KANIZSA 1979, TAMPIERI 1956, ATTNEAVE 1977, ARNHEIM 1969 p. 86) fully appreciated the significance of this aspect of amodal perception.

Conclusion

The model presented here represents a preliminary attempt to express the components of visual perception in terms that can be incorporated in a quantitative model of subjective experience. Many of the aspects of the model, such as the volumetric perception of depth, the boundedness of spatial perception, the rotation of the phenomenal world, amodal perception, and perception outside the visual field, reflect properties of perception that were identified decades ago by the Gestaltist. However these aspects of perception have received little attention in more recent decades. The reason for this oversight is that these properties are not easily expressed in the neural network paradigm that has come to dominate the description of perceptual phenomena in psychology. This has led to a growing gap between models of spatial perception and the subjective experience of the visual world. In 1935 Kurt KOFFKA wrote: *"American psychology all too often makes no attempt to look naively, without bias, at the facts of direct experience, with the result that American experiments quite often are futile. In reality experimenting and observing must go hand in hand. A good description of a phenomenon may by itself rule out a number of theories. ... Without describing the environmental field we should not know what we had to explain."* (KOFFKA 1935, p. 73). This statement remains as true today as it was six decades ago.

Summary

The Gestalt principle of isomorphism suggests that the subjective experience of spatial perception is a valid source of evidence for the nature of the underlying neurophysiological representation. A perceptual modeling approach is proposed, to quantify the information manifest in spatial perception without regard to the neural mechanism by which that information is encoded. This approach highlights aspects of perception commonly overlooked by neural network models, including the volumetric nature of spatial perception, the perception of empty space around perceived objects, the bounds of perceptual space, the rotation invariance of perceptual representation, amodal perception, and perception outside of the visual field.

Zusammenfassung

Das Isomorphismus-Konzept der Gestalttheorie weist darauf hin, daß die phänomenale Raumwahrnehmung eine zuverlässige Basis für die Natur der zugrunde liegenden neurophysiologischen Repräsentanz ist. In diesem Beitrag wird ein Modellierungsansatz vorgeschlagen, die in der Raumwahrnehmung manifeste Information zu quantifizieren - und zwar unabhängig von den neuronalen

Mechanismen durch welche die Information codiert wird. Dieser Ansatz betont Aspekte der Wahrnehmung, die üblicherweise von neuronalen Netzwerk-Modellen übersehen werden - wozu die spezifische Raum-Metrik der Raumwahrnehmung, die Wahrnehmung eines leeren Raumes um wahrgenommene Objekte, die Begrenzungen des Wahrnehmungsraumes, die Rotations-Invarianz der Wahrnehmungsrepräsentation, amodale Wahrnehmung und letztlich Wahrnehmung außerhalb des visuellen Feldes gehören.

References

ARNHEIM, R. (1969). *Visual Thinking*. Berkeley, University of California Press.

ATTNEAVE, F. (1954). Some Informational Aspects of Visual Perception. *Psychology Reviews, 61,* 183-193.

ATTNEAVE, F. (1977). The Visual World Behind the Head. *American Journal of Psychology, 90(4),* 549-563.

BIEDERMAN, I. (1987). "*Recognition-by-Components: A Theory of Human Image Understanding*". *Psychological Review, 94*, 115-147.

BLUMENFELD, W. (1913). Untersuchungen Über die Scheinbare Grösse im Sehraume. *Z. Psychol., 65,* 241-404.

BLANK, A. A. 1958 Analysis of Experiments in Binocular Space Perception. *J. Opt. Soc. Amer., 48,* 911-925.

BORING E. G. (1933). *The Physical Dimensions of Consciousness*. New York: Century.

BROAD, C. D. (1978). *Kant - an introduction.* Cambridge: Cambridge University Press.

CHARNWOOD J. R. B. (1951). *Essay on Binocular Vision*. London, Halton Press.

COREN, S. WARD, L. M. & ENNS J. J. (1979). *Sensation and Perception*. Ft Worth TX, Harcourt Brace.

FOLEY, J. M. (1978). Primary Distance Perception. In R. Held, H. W. Leibowitz, & HJ. L. Tauber (Eds.), *Handbook of Sensory Physiology*, Vol VII Perception (pp. 181- 213). Berlin: Springer Verlag.

GALLI, A. (1932). Über mittels verschiedener Sinnesreize erweckte Wahrnehmung von Scheinbewegung. *Arch. f. d. Ges. Psych., 85,* 137-180.

GIBSON, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

GRAHAM, C. H. (1965). Visual Space Perception. In C. H. Graham (Ed.). *Vision and Visual Perception* (pp. 504-547). New York: John Wiley.

GROSSBERG, S. (1987a). Cortical dynamics of three-dimensional form, color and brightness perception. I. Monocular theory. *Perception & Psychophysics, 41,* 87-116.

GROSSBERG, S. (1987b). Cortical dynamics of three-dimensional form, color and brightness perception. II. Binocular theory. *Perception & Psychophysics, 41,* 117-158.

HEELAN, P. A. (1983). *Space Perception and the Philosophy of Science* Berkeley: University of California Press.

HELMHOLTZ, H. (1925). Physiological Optics. *Optical Society of America, 3,* 318.

HILLEBRAND, F. (1902) Theorie der Scheinbaren Grösse bei Binocularem Sehen. *Denkschr. Acad. Wiss. Wien (Math. Nat. Kl.)., 72,* 255-307.

HUBEL, D. (1988). "*Eye, Brain, and Vision*". New York: Scientific American Library.

JULESZ B. (1971). *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.

KANIZSA, G. (1979). *Organization in Vision.* New York, Praeger.

KANT, I. (1781). *Critique of Pure Reason.*

KAUFMAN (1974). *Sight and Mind.* New York: Oxford University Press.

KOENDERINK, J. & Van DOORN A. (1976). The singularities of the visual mapping. *Biological Cybernetics 24,* 51-59.

KOENDERINK, J. & Van DOORN A. (1980). Photometric invariants related to solid shape. *Optica Acta 27,* 981-996.

KOENDERINK, J. & Van DOORN A. (1982). The shape of smooth objects and the way contours end. *Perception, 11,* 129-137.

KOFFKA, K. (1935).. *Principles of Gestalt Psychology*. New York: Harcourt Brace & Co.

KÖHLER, W. (1938). *The Place of Value in a World of Facts.* New York: Liveright.

KÖHLER, W. (1947). *Gestalt Psychology*. New York: Liveright.

KÖHLER, W. & HELD R. (1947). The Cortical Correlate of Pattern Vision. *Science, 110.* 414-419.

KÖHLER, W. (1929). Ein altes Scheinproblem. *Die Naturwissenschaften 17,* 395-401. Reprinted in Henle M. (Ed.) (1971). *The Selected Papers of Wolfgang Köhler.* New York, Liveright.

LEHAR, S. & McLOUGHLIN, N. (1998). Gestalt Isomorphism II: The Interaction Between Brightness Perception and Three-Dimensional Form. *Perception (submitted for publication).*

LUNEBURG, R. K. (1950). The Metric of Binocular Visual Space. *J. Opt. Soc. Amer., 40,* 627-642.

MARR D. & POGGIO T. (1976). Cooperative Computation of Stereo Disparity. *Science, 194,* 283-287.

MARR, D, (1982). *Vision.* New York: W. H. Freeman..

McLOUGHLIN, N. & GROSSBERG, S. (1998). Cortical Computation of Stereo Disparity. *Vision Research, 38,* 91-99.

MÜLLER G. E. (1896). Zur Psychophysik der Gesichtsempfindungen. *Zts. f. Psych. 10.*

O'REGAN, K. J., (1992). Solving the 'Real' Mysteries of Visual Perception: The World as an Outside Memory *Canadian Journal of Psychology, 46,* 461-488.

PINKER, S. (1984). Visual Cognition: An Introduction. *Cognition, 18,* 1-63.

REED E. S. (1988). *James J. Gibson and the Psychology of Perception.* New Haven CT: Yale University Press.

TAMPIERI, G. 1956 Sul Completamento Amodale di Rappresentazioni Prospettiche di Solidi GeometriciSS. *Atli dell' XI Congresso Degli Psicologi Italiani*, ed. L. Ancona, pp 1-3 Milano: Vita e Pensiero.

TODD, J, & REICHEL, F, (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces *Psychological Review, 96,* 643-657.

Anschrift des Verfassers:

Steven Lehar

14 Crooked Lane

Manchester MA 01944

USA

e-mail: slehar@sharp.bu.edu