Vision Research 48 (2008) 2128–2142

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

# Illusions in the spatial sense of the eye: Geometrical–optical illusions and the neural representation of space

Gerald Westheimer *

*Division of Neurobiology, University of California, Berkeley, CA 94720-3200, USA*

ARTICLE INFO

*Article history:*
Received 29 February 2008
Received in revised form 20 May 2008

*Keywords:*
Metrical properties of visual space
Contours
Spatial primitives
Perceptual rules
Hierarchical order

ABSTRACT

Differences between the geometrical properties of simple configurations and their visual percept are called geometrical–optical illusions. They can be differentiated from illusions in the brightness or color domains, from ambiguous figures and impossible objects, from trompe l'oeil and perspective drawing with perfectly valid views, and from illusory contours. They were discovered independently by several scientists in a short time span in the 1850's. The clear distinction between object and visual space that they imply allows the question to be raised whether the transformation between the two spaces can be productively investigated in terms of differential geometry and metrical properties. Perceptual insight and psychophysical research prepares the ground for investigation of the neural representation of space but, because visual attributes are processed separately in parallel, one looks in vain for a neural map that is isomorphic with object space or even with individual forms it contains. Geometrical–optical illusions help reveal parsing rules for sensory signals by showing how conflicts are resolved when there is mismatch in the output of the processing modules for various primitives as a perceptual pattern's unitary structure is assembled. They point to a hierarchical ordering of spatial primitives: cardinal directions and explicit contours predominate over oblique orientation and implicit contours (Poggendorff illusion); rectilinearity yields to continuity (Hering illusion), point position and line length to contour orientation (Ponzo). Hence the geometrical–optical illusions show promise as analytical tools in unraveling neural processing in vision.



## 1. Introduction

Illusions are situations when a percept differs from the generating stimulus in a meaningful yet misleading way. Aside from serving as warnings of the untrustworthiness of our senses, their scientific interest lies in the insight they allow into the sorting of sensory signals in the process of generating percepts.

The study of illusions goes through several stages. First, they have to be identified and described ("What?"). Then there are two other stages whose priority depends on the scholarly cast of the enquirer. A psychologist or evolutionary biologist will want to know "Why?" and the physiologically inclined, "How?" At the root of the former is the proposition that an illusory perceptual discrepancy reveals underlying imperatives of an organism's exploration of its environment, whereas the latter regards them as windows into the operation of the neural apparatus that parses sensory information.

Researchers depend on the state of science in their time. When the mechanism of the eye was barely understood and that of the inner workings of the retina and brain rudimentary at best, the student of visual illusions tended to concentrate on discovering and describing the phenomena. Because they are numerous, this proved to be an exceedingly productive enterprise. The direction of more searching analyses is predicated on the depth to which they can be taken. Early attempts to arrive at principles of perceptual organization by the Gestalt School in the first third of the 20th century were not sufficiently successful to fold in visual illusions as guides. In the current climate of enormous expansion of knowledge of the nervous system, it is natural to try to interdigitate with it. Detailed search of neural structure and function for hints to elucidate illusions may at this juncture appear more productive, but in the end cannot substitute for behavioral research. Hering's stance vis-à-vis Helmholtz need always be kept in mind: To know what a clock is all about, it helps, while taking it apart and looking at the cogs and wheels, also to examine its face. Few can match Hering's genius to do that effectively, so in most instances we follow Helmholtz's example and proceed to a methodical analysis of illusions through the sequence of physical, anatomical and physiological stages in the elaboration of the percept from the stimulus, being fortunate in having available the powerful and uniquely applicable tool of psychophysics.

Viewed from this vantage, the list of illusions in the spatial sense of the eye can at the outset be pared down by excluding

* Fax: +1 510 643 6791.
*E-mail address:* gwestheimer@berkeley.edu

0042-6989/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2008.05.016

some phenomena that, strictly speaking, are not illusions at all. The Necker cube (Fig. 1), sometimes called a reversible figure, is an inherently ambiguous pattern. It allows more than one interpretation, but only one at a time and neither is misleading. Figures of the Penrose (Penrose & Penrose, 1958) (Fig. 1) and Escher type, require detailed parsing of the figure's components and deductive reasoning to reveal the impossibility of their physical realization; their class differs from those in which the misleading geometrical properties are instantly, or as it is sometimes called, pre-attentively, evident. The distinction between the two modes of observation seems intuitive when phrased in this manner but can in fact be demonstrated operationally. In the German literature from Kant to Hilbert (Hilbert & Cohn-Vossen, 1932), the word used for the act of immediate, unmediated perception, antecedent to dissection by reason, is *Anschauung*, often translated as apperception.

It once was thought that the dichotomy between pre-attentive and more deliberate judgments allowed a separation of layers of neural processing, for example the primary visual cortex, sometimes referred to as ''early,'' *versus* ''higher'' cortical levels. But this kind of divide ignores the prominent connectivity––invariably in both directions!––within the cerebral cortex and the evidence that has been accumulating for top-down conditioning, due to signals originating further upstream, of neural circuits right at the entry points of the sensory input.

For the sake of concision, the discussion here also excludes the following phenomena that fit only peripherally into the framework:

*Perspective drawings*, *trompes l'oeil* and other situations in which perfectly valid and veridical monocular views evince conflicts when non-matching binocular disparity clues are added; they might be classed more appropriately under the heading of perceptual dissonance rather than visual illusion.

*Illusory Contours*, i.e., contours that are not explicitly drawn but only, so to speak, hinted at; they behave in many respects like actual contours of low contrast (Petry & Meyer, 1987). Figures can be generated with them and they can participate, if less than optimally, in the more overt geometrical illusions, and even elicit responses in some neurons in the visual cortex (von der Heydt, Peterhans, & Baumgartner, 1984). Though they are illusory effects in the spatial domain, their interest lies more in how contours are generated in the visual system than how, once generated, the contours' spatial properties, and indeed visual space itself, emerge. Hence they will not be given detailed consideration in this study.

*Time* is separable from space as a category in the human experience. This is not to say that there is no interaction between the two, even after leaving aside the singular perception of movement, which surely involves both. This study is focused on space, but insofar as empirical findings enter, exposure duration can matter and so can asynchronous display of components of configurations. Of relevance are the so-called figural aftereffects (Köhler & Wallach, 1944) in which some figures, or their components, when shown first, influence the perception of the geometrical properties of what is shown subsequently. In general, the phenomena are closely related, and often identical with, what is seen with synchronous exposure. The time course of decay of such aftereffects is usually of the order of a very few minutes (Hammer, 1949). Explanations are most often offered in terms of ''adaptation,'' where the neural state lingers after the extinction of the immediate stimulus and continues to operate in the same manner but with gradually diminishing intensity. Figural aftereffects and simultaneous spatial interaction of figures or their components may seem to be quite different phenomena, but from the perspective of modern neurophysiology subsuming the two under the same heading would not present a problem.

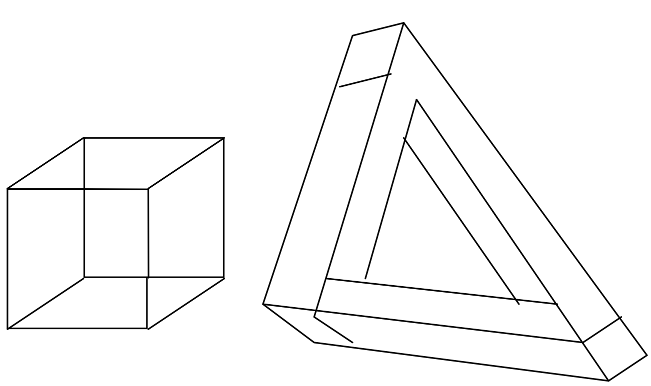

**Fig. 1.** Necker cube, a reversible figure, (left) and Penroses' impossible object (right) are spatial configurations that are not geometrical-optical illusions—the Necker cube is ambiguous not illusory, and the geometrical properties of the various section of the Penrose triangle are each by themselves perfectly valid.

*The moon illusion* has been discussed in the literature for centuries. From its earliest descriptions, it has been identified as having a deeply cognitive origin, since it is regarded as depending on intrinsic knowledge of absolute size and of the law of size constancy relating the absolute distance of an object and the angle it subtends at the eye (Ross & Plug, 2002). Once it is understood that absolute size of the retinal image is a pointer to the absolute distance at which a familiar object is located, and once there is a rudimentary appreciation of the absolute size of the retinal image, the relative distances of objects can be estimated. Hence the moon's disk, when seen in conjunction with familiar objects of known size, will be taken as implying a larger distance and therefore a larger absolute diameter than when it is high in the sky and seen in isolation. This interpretation, on which there is more or less universal agreement, is contingent both on the ability to gauge the absolute size of a retinal image and of its engendering object, and on the available knowledge of the relationship between size and distance. Its ontogeny is therefore a critical issue (Leibowitz & Judisch, 1967; Wohlwill, 1960).

## 2. Two worlds: objects and visual experience

In approaching illusions, two concepts are involved: the world of objects and that of an observer's visual experience. The two realms, the eye's object space and the manifold housing visual percepts, are separate and quite distinct.

The stimulus situation in the real world of visual objects is unproblematic: it is three-dimensional and Euclidean. The physical qualities of objects, including their location and geometrical properties, can be specified with arbitrary precision and it is understood that this has been done prior to any attempt at assigning a visual illusion to low-level neural factors or to higher-order cognitive ones. In addition, it is mandatory to be assured that there are not explanations in terms of optical imaging in the eye. These days it is a matter of routine to define the light distribution on the retina, though the position of the observer's eyes matters of course.

If the real, physical world is one end of the arc in the scientific study of visual illusions, the other is that of the observer's percepts. Though percepts are, in their essence, subjective phenomena, public knowledge of them being based on reports of observers' visual experiences, they need not be regarded as nebulous entities. To the contrary, they may be characterized with much the same rigor that is customary in most branches of science. Good analytical tools are available for describing, in adequate detail, the representation of visual stimuli within the organism:

(a) the report of the visual experience by the observer;
(b) quantitative characterization of the imputed visual experience by manipulation of stimuli and observation of the associated behavior in an experimental subject, human or animal; and
(c) measurement of neural activity recorded in an experimental subject, human or animal, allowing some conclusion about on object's representation in the nervous system.

To use a specific example: suppose we are interested in the illusion where a straight line looks curved when an adjoining circle is added to the display. Under (a) above, the observer would report this experience. Under (b) an animal trained, or human observer instructed, to respond differently to straight and curved lines reveals a change from "straight" to "curved" when the contextual circle is introduced. A particularly effective means of investigation is the "nulling" experiment in which the test line is given the opposite curvature required to make it appear straight. Under (c) above, neural units or brain images might reveal quite specific differences for straight and curved line stimuli, and exhibit a telling change from one to the other with the application of the circle.

Whereas the word visual illusion unquestionable fits (a) and with little extension also (b), it would be more appropriate to speak in situation (c) of the "neural substrate of the illusion."

In what follows no attempt will be made to substitute for the excellent treatises enumerating these illusions and the many experimental attempts to solidify the theoretical formulations to account for them (Coren & Girgus, 1978; Morgan, Hole, & Glennerster, 1990; Robinson, 1998; Wade, 1982). Routinely more than one cause is found to be responsible for the deviation of a percept from its generating stimulus. The most economical approach then is a sequential one, proceeding from known and specifiable effects to less securely established and more conjectural ones. In this review this sequence starts with the physics of the stimulus situation and the retinal image, and then goes on to neural processing in the retina and the putative ones in the primary visual cortex. It is not implied that these are the more important stages; it is just that phenomena anchored in optics and anatomy can be identified more securely and with more general applicability. The essentially subjective element in visual illusions can never be ignored.

## 3. Illusory spatial effects caused by brightness influences

In vision, misleading differences between objects and their corresponding percepts can occur in many stimulus attributes; this review concentrates on the spatial illusions, those of the location and shape of configurations, and leaves aside the ones involving the intensive attributes, light and color. However, the light sense and the spatial sense cannot easily be separated. Except on the rare occasion in the use of a Ganzfeld, the study of how we see brightness requires that stimuli be identifiable and hence spatially circumscribed. Conversely, in examining the spatial sense, locations and extent have to be indicated, and that can only be achieved by markers laid out by brightness or color differences.

Ignoring some effects caused by passive optical factors in the refractive apparatus of the eye, e.g., chromastereopsis, illusions in the spatial sense of the eye can conveniently be divided into those in which spatial deformations are a consequence of the exigencies of the processing in the domain of brightness, and the true geometrical–optical illusions, which are misperceptions of geometrical properties of contours in simple figures.

In the imaging by the eye's optics and generation of neural signals in the retina, activity is always spread across a small region of the retina. When even a single luminous point such as a star, the most elemental spatial stimulus, is presented to the eye, the retinal response will not be punctate. The basics of the light distribution are summarized in the well-known point-spread function (Fig. 2) and, because all stimuli can be regarded as a collection of points, in any situation one can specify where and how much light is delivered to the retinal receptors (Westheimer, 2006). They in turn

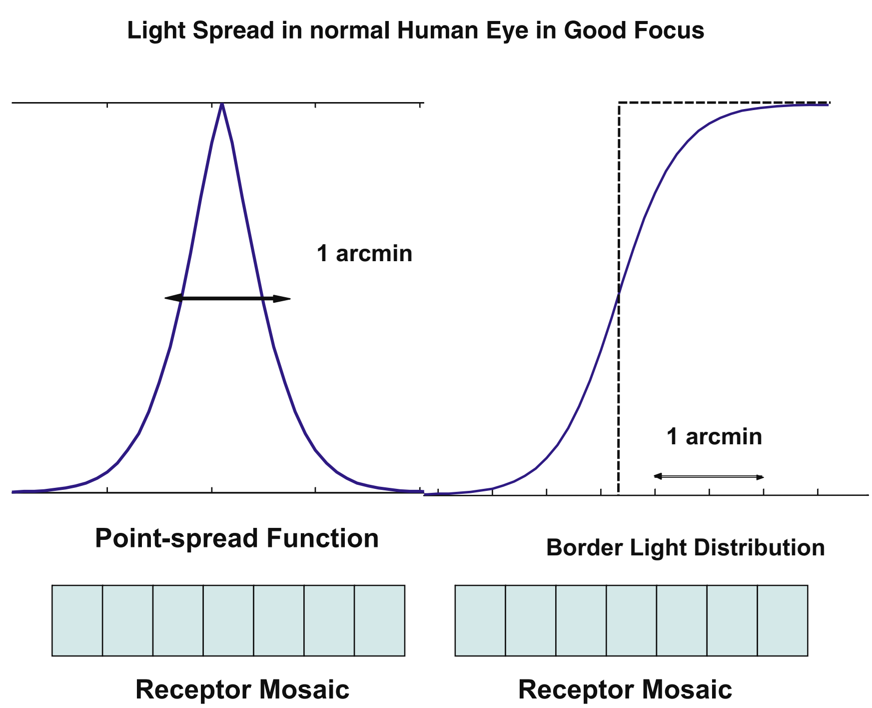


**Fig. 2.** Left: spread of light in the retinal image of a point target in the visual field. All objects can be thought of as a collection of points and their retinal image distribution is the superposition of that of all such points. As an example, the light distribution in the image of an edge is shown on right. The axis of ordinates is in relative light units, that of the abscissas in minutes of arc, which is of the order of the resolution limit in best foveal vision.

communicate through a complex circuit to the ganglion cells, whose activity can be quite well described in most instances. Hence the information about the outside world that is communicated by the eye to the brain is in its broadest form sufficiently understood to state whether, and if so, to what extent, optical and retinal stages are culpable in the mischaracterization of a stimulus.

The most prominent of the spatial illusions secondary to brightness effects is the phenomenon of irradiation whose study goes back hundreds of years when it was first recognized that the brighter a star the larger it appears. Goethe, who pioneered the study of describing visual experiences in a variety of settings of light, dark and color, quoted Kepler as reporting this observation (Goethe, 1808). Certain illusions involving border mislocation similarly contain elements of irradiation (Pierce, 1901).

### 3.1. Star brightness and apparent size

If pupil size is not an issue, the shape of the distribution of light in the retinal image of a star remains invariant with increase in its intensity; it is just scaled up. If bright stars look larger than dim ones, the cause would have to be other than purely optical. In accord with our program to try for low-level explanations first, the next question concerns retinal processing. The transduction of light takes place in the receptor cells, the rods and cones, each acting as a single compartment, but only in the foveal cones is the grain of the individual members of the receptor mosaic conserved in the forward passage into further neural stages. Data have not been accumulated, but it is likely that the irradiation effect of apparent star diameter is confined to foveal vision. Because the point-spread function covers at a minimum perhaps a dozen cones, more and more of them will reach and then exceed their threshold when the source intensity is increased. If indeed an observer can distinguish whether an image is one, three or five cones in diameter, then a purely optical and anatomical explanation of the phenomenon is available. This interpretation is helped by a non-linear stage in the elaboration of the brightness signal in the outer retina, where it is still graded and has not yet been transformed into action potentials. It is the Naka–Rushton compressive non-linearity, first described in the fish (Naka & Rushton, 1966) but more recently found also in mammalian eyes (Valeton & van Norren, 1983). In its operation (Fig. 3) a fixed ratio between two input light levels results in a smaller output ratio the higher the intensity. Processing by such a circuit of a series of distributions of equal shape but increasing amplitude will generate output distributions with smaller increases in height but progressive changes in width. Fig. 4 depicts the results when the retinal image of a star at 4 levels of magnitude, each one log unit apart, has been processed by a circuit with a Naka–Rushton compressive non-linearity with suitable parameters. For a better appreciation of the width changes with intensity, the "neural spread functions" have been normalized.

Retinal anatomy at this distance scale requires, of course, description in terms of individual cells and this is as yet a formidable experimental proposition. Hence the measure of the distributions' width at half-height can be regarded only as a hint of the lateral spread of neural excitation. Still, it is evident that a model of retinal filtering incorporating optical light spread and a compressive non-linearity in the intensity dimension can account for the increase in seen width of a star with its brightness. The assumption that these changes will survive all the succeeding stages of retinal and retino-cortical transmission is not unreasonable.

### 3.2. White areas look larger than equal-sized black areas

Related to the apparent enlargement of stars with increase in their intensity and also subsumed under irradiation is the well-documented phenomenon of bright spaces looking larger than dark ones of identical physical dimensions. It is said to have been understood by the architects of Greek temples.

Its origin lies in the process by which a border, i.e., an abrupt change in luminance, is assigned location. For its understanding it is necessary to examine the optical image on the retina of a border. It has the shape of a symmetrical ogive (Fig. 2, right) whose width may change with changes in quality of the imagery, and whose overall height increases with target luminance level but

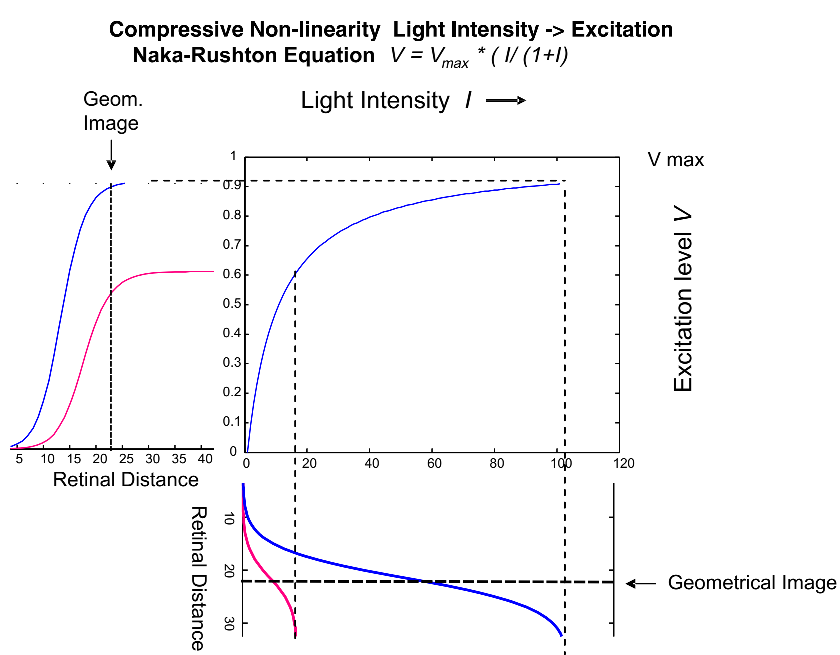


**Fig. 3.** When optical light signals from black/white edges of different amplitude (bottom, oriented sideways) are processed by retinal circuits, they undergo a compressive non-linearity and generate neural signals that are transformed and whose position is displaced relative to the location of the geometrical edge (left of figure).

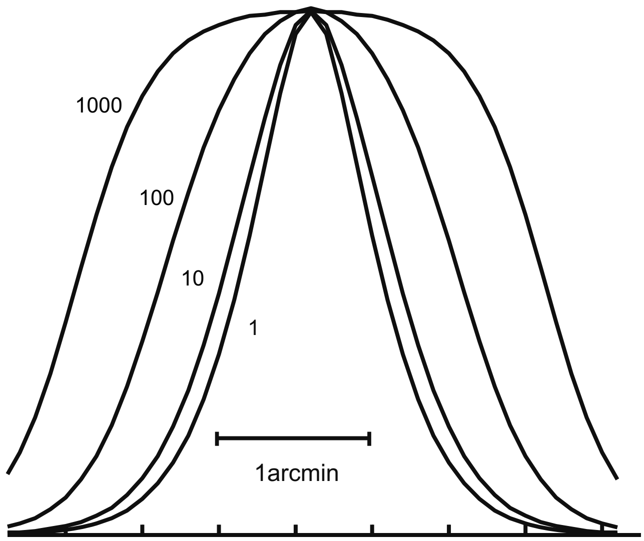


**Fig. 4.** "Neural spread distributions" of a star at four levels of magnitude as they emerge after having been subjected to optical spread of light and neural processing though a Naka–Rushton non-linearity. Signal compression increases progressively with light intensity, effectively widening the distribution the brighter the star. For ease of comparison, the four curves are shown normalized at their peak.

whose midpoint coincides with the edge of the hypothesized ideal geometrical image. However, measurements of the perceived location of the border show that it is not at the midpoint but rather about 0.5 arcmin towards the dark side. A convenient explanation here also is the compressive non-linearity in the generation of the brightness signal. When the symmetrical ogive is subjected to such a transformation, the resultant edge distribution is no longer symmetrical but has its points of half-height as well as of inflection shifted to the dark side (Fig. 3) and, moreover, the more so the higher the retinal illuminance of the bright region. A model utilizing the known eye's light spread and the Naka–Rushton transformation with a suitable parameter yields values that give a good fit of the observed border shifts.

### *3.3. Shifted-chessboard and other illusions with components of border mislocation*

In a class of visual illusions, of which the one called shifted-chessboard is the most well known, offset black and white borders produce apparent tilts of dividing lines. Subjected to thorough analysis quite soon after their discovery, they depend critically on processing of brightness. A very elaborate member of the class is the Kitaoka pattern (Kitaoka, 1998) where, Fig. 5, the illusion disappears when the black and white panels are replaced by isoluminant heterochromatic ones. This is evidence that at least part of the illusion is due to the processing of brightness. There are many variants of this effect and they may not all be covered by the same explanatory principles.

The shifted-chessboard pattern has recently been subjected to a detailed deconstruction which reveals it has at least four components (Westheimer, 2007). At the beginning is the displacement of the black/white border, which has just been discussed, and is the consequence of the optical light spread in the eye and the compressive non-linearity of brightness processing in the inner retina. Neither, however, can account for crimping of corners because this is evident in black corners on a white background and as well as the reverse. But when another property of retinal circuitry, the center-surround organization is folded in, much more of the shifted-chessboard pattern deformations can be explained and actually nulled out experimentally. However, some of the effect still remains even then, illustrating the multiple origin of these illusions as well as the reason for the strategy of trying to associate effects sequentially to optical, retinal and then cortical factors. What remains after this has been accomplished are deformations of the configurations' geometrical properties in a stricter sense. They survive after the factors ascribed to optical imaging and retinal preprocessing have been eliminated, leaving just the contours for the operation of the geometrical–optical illusions, to which we now turn.

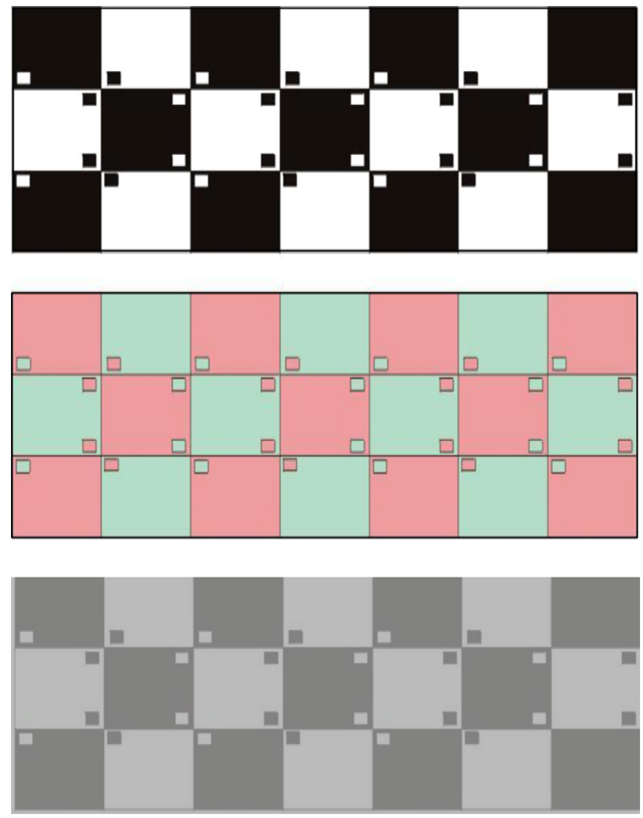

**Fig. 5.** An illusion designed by Kitaoka (1998), a variant of the shifted-chessboard illusion of Münsterberg, illustrates how heterochromatic isoluminance can help in distinguishing between a true geometrical–optical illusion and an illusion that has its origin in the processing of brightness and color.

## 4. Geometrical–optical illusions

The distinguishing feature of these illusions is that they relate to misjudgments of geometrical properties of contours and are quite robust to contrast as well as to contrast polarity, that is, they show up equally for dark configurations on a bright background and the reverse. It does not necessarily follow that the phenomena remain unaffected when the contrast polarity or chromaticity within a single configuration is mixed. Therefore, for their investigation it is best if the contours are sharply delineated and this is most effectively achieved by showing them in black against a white background or vice versa.

A historical survey of the discovery of geometrical–optical illusions is included in Appendix I. Once established, the topic became very popular and the number of illusions fitting the type has grown astronomically. Excellent enumerations are available in the compendiums already mentioned (Coren & Girgus, 1978; Robinson, 1998; Wade, 1982) and it is not the intention here to duplicate or update their efforts, even if this were a feasible proposition. Because the issues involved in perception are most often very complex, it is worth trying to delineate the class of spatial illusions that relate to the discipline of geometry and decouple them from the much wider topic of object perception. Even the word "shape"––as in shape constancy––can have connotations well beyond what can be discussed under the heading of geometry, broadly conceived. And the more cognitive the content of a topic, the more likely it is that there will be misjudgments and hence situations in which the word illusions might properly be applied. Before proceeding further it is, therefore, incumbent on us to try for a comprehensive definition of the geometrical part of the term geometrical–optical illusion. Central to the whole discussion is the question of how to approach, both qualitatively and quantitatively, the nature of the lay-out of contours as they appear to us.

## 5. Geometry and visual space

One of the tasks of the scientific discipline of visual perception is to seek a satisfying description of the mapping of the physical object space of visual stimuli onto visual space. There is a one-to-one correspondence between the two––the situations covered in this review exclude hallucinations and scotomas. By and large the relative location of objects is conserved in the mapping, though not universally. Broadly defined, geometrical–optical illusions are situations in which there is an awareness of a mismatch of geometrical properties between an item in object space and its associated percept.

Although it can be accessed only indirectly, the private world of an observer's visual experiences, containing perceived objects arrayed in their different locations, is undoubtedly a space according to all definitions of this word. This visual space has, like the physical object space, three dimensions, but it cannot immediately be taken for granted that the two are alike in all other respects. With some reservations and modifications, many of the techniques of studying the nature of spaces in general can, however, be utilized enabling us to call on a rich tradition of scholarship in geometry for help in delineating the relationship between physical object and visual space. The mapping of the content of one space into another is a transformation and this inquiry centers on the question whether geometrical–optical illusions could possibly have their roots in laws governing the transformation between Euclidean object space and visual space.

The broadest qualitative questions are the deepest because they border on imponderables such as what is a dimension of space or what is a straight line. Geometers can invent constructs and their rules of manipulation, but in the end they still need to fall back on givens––the axioms, whose roots are actually in human perception. The challenge here is not to go full circle, but to attempt a cohesive framework that encompasses perception as well as the axiomatics and workings of geometry as a discipline. This necessitates leaving aside the more abstract kind of spaces that mathematicians have invented, and also some varieties of geometry that are obviously unsuitable to the current discourse, topology, for example, which sees no difference between a cube and a sphere.

A helpful approach is the influential one of Felix Klein in his "Erlanger Programm" of 1872. Klein starts with defining a space as a manifold containing structures, and this is evidently applicable to both the space of physical objects and that of visual percepts. According to Klein, in such a manifold containing structures, the geometry is defined by the group of transformations that allows the structures to remain invariant. Hence for the specification of a geometry, both the acceptable group of transformations and what is regarded as invariant have to be stated. For example, the transformations considered may be just simple translations and rotations. Structures are, at the outset, just lines and figures like squares and triangles. An attempt is made in Appendix II to formulate an approach to geometrical–optical illusions as manifestations of metrical properties of the observer's visual space. In principle this is an inviting concept and, as has been demonstrated when applied to color vision, if carefully framed need not suffer from internal contradictions. But incorporating the many empirical observations that necessarily have to be accommodated makes demands that at this stage are not readily surmountable.

In taking over into perception the concepts developed in geometry, very difficult problems arise at the very outset. These are much more penetrating than merely rejecting the failure to differentiate between cubes and spheres in topology. Mach already pointed out that observers would report a square as unchanged when displaced in the fronto-parallel plane, but not when it is rotated through 45°; a diamond's geometrical properties (Fig. 6) may be identical with a square's within the definition of Klein's principal group, but this does not necessarily extend to the immediate perception. There is subtle distinction between the instant unfiltered report of a percept and the result of a reasoned analysis based on logical deduction from premises external to the current situation. Helmholtz, quintessential physicalist that he was, was unwilling to make that distinction and conflated the two by his proposition of "unconscious inference." But if one accepts Mach's observation, then by Klein's definition––a geometry is defined by the properties of structures that remain invariant with the principal group of transformations––a space wherein a diamond is something different than a square of equal physical dimension belongs to a family well beyond the confines of what geometers have considered. This opens up the challenging question of how many kinds of visual space there are and how to navigate between them, in particular since they are all only indirectly accessible yet should be capable of characterization by some operations. Rather than embarking on an agonizing debate, it helps to defuse arguments by limiting consideration to situations amenable to experimental programs in which stimuli are confined to an operationally definable set and responses to simple value- and preconception-free categories: Which of the two patterns is larger? Are specific features of patterns aligned or not? Is the contour straight or not? This leaves a vast perceptual terrain wide open, but at least allows the more circumscribed one of the geometrical–optical illusions to be charted.

The discussion of geometrical–optical illusions becomes a lot easier if consideration is restricted to a Euclidean object plane of two dimensions. For areas extending only 10° or 20° of arc in the fronto-parallel plane, this is surely true and objects can be specified on a Cartesian coordinate system centered on the fixation point.

The third spatial dimension is encoded by the peripheral visual apparatus as a disparity, i.e., a difference between retinal images of

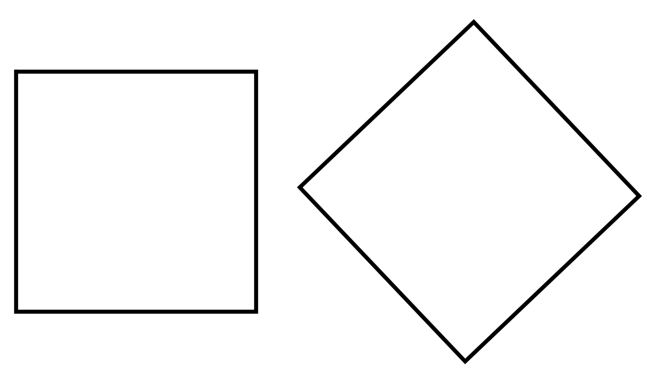

**Fig. 6.** Mach pointed out that an observer will not immediately perceive a diamond (right) as being the same configuration as a square (left) even though both have identical properties as defined within all types of geometry. To allow visual space to have this kind of property would mean that it is not amenable to analysis within the framework of geometrical reasoning.

the two eyes. It is utilized by a sophisticated neural machinery and produces a sensation quite separate from that arising from the two-dimensional spatial coordinates that project the three-dimensional outside world onto the two-dimensional retina by way of lines of sight and visual angles. The sensing of binocular disparity, folded in with the appreciation of depth from perspective and secondary clues such as interposition, yields a compelling three-dimensional visual space. Fortunately, the fundamentals of geometrical–optical illusions, most of which appear on inspection of figures drawn on a sheet of paper, can be studied without consideration of the third dimension in the realms of object space, of the retinal imagery and of the associated neural circuitry and observers' perceptual experience.

It is necessary in a brief aside to separate out some clinical situation in which a patient sees distortions of the straight lines of a rectangular grid in object space due to pathologically-caused disarray of the retinal mosaic. The location label intrinsic to each retinal afferent and its connection with and integration into the neural visual stream is retained, but relative positions with respect to the object have been mechanically disturbed unbeknownst to the next visual stages, preventing compensation through plastic changes in the cortex.

A faithful mapping implies that when two configurations, or components of configurations, are identical in object space, their counterparts in perception are also identical. Failure to achieve this is evident at many levels of vision; geometrical–optical illusions deal specifically with failure in such basic properties of contours as curvature or straightness, their orientation, and in the magnitude of distances and angles. All of these can, of course, be precisely established in object space, but with suitably designed experiments and full explanation of criteria there is little impediment to their characterization also in visual space. The preconditions for scientific measurements of the transformation between the two have therefore been met. On the other hand, attempts at more detailed categorization of illusions, e.g., of size or alignment, have not been so successful; they are generally based on what properties might be grouped together. This can, however, create difficulties, for example when illusory differences in magnitude of angles manifest themselves in illusory differences in alignment, which in turn might be ascribed to illusory failure of rectilinearity. Actually, there may even be simpler underlying principles, such as a vertical/horizontal anisotropy or one of oblique compared to cardinal orientation of contours. In pursuing these questions of classification, it is desirable at the outset to retain a fluid stance, because of necessity any system of classification will involve externally imposed partitions. To many, the most tangible ones are derived by looking for clues inside the nervous system.

## 6. The neural representation of space

The spectacular progress made in the examination of internal neural mechanisms now makes it possible to proceed with a study of geometrical–optical illusion by side-by-side consideration of physiological and psychophysical aspects. Basic to the former is the question of how the spatial attributes of visual stimuli are encoded in the nervous system, but this opens up a line of inquiry that is straight-forward only in its beginning stages. Optical imagery by the eye's refractive apparatus and the anatomical structure of the retina insure a reasonably faithful point-for-point reconstitution of the outside world at the level of the receptive layer. Compartmentalized individual retinal receptor cells, only marginally compromised by cell-to-cell electrical couplings, have their mosaic remapped onto ganglion cells; so, at a minimum, the number of individual local signatures in a retinotopic map in the primate is of the order of $10^6$ and this number provides an adequate depiction of the spatial dissection carried out by the visual stages prior to the cortex. There is some reorganization, center/surround antagonistic coupling, for example, intended to highlight differences, but that does not necessarily reduce the grain size of the retinotopic map which, to a good approximation, is preserved within the path into the visual cortex. That is to say, in the lateral geniculate nucleus and the primary visual cortex signals from neighboring points in the visual field remain neighbors, though the magnification scale changes from the fovea to the periphery of the visual field (Dow, Snyder, Vautin, & Bauer, 1981). Viewed externally, the lay-out of the visual field on the cortical surface is deformed (Schwartz, 1977), but that does not by any means imply that spatial functional relationships of neural signals have experienced any distortions. For example, stimuli arising in three points in the visual field marking out a triangle will produce a homologous pattern of cortical excitation. This would seem to be an obvious statement, but it serves to dispose of theories of radical reorganization, such as representation in the form than spatial frequency rather than spatial location.

Isomorphism––exact matching of spatial properties of object patterns in their corresponding cortical signals––fails, however, quite soon and in several ways, and not just because of inevitable deformations due to the anatomical dictates of fitting the cortex into the skull:

### 6.1. Overlaying of representation of several spatial parameters

From the first findings (Hubel & Wiesel, 1959), it is known that neural elements in the visual cortex are selective for contour orientation, and length. Hence, while broadly speaking there is a retinotopic map in the primary visual cortex, individual neurons respond to (and presumably are coded for) spatial attributes that encroach on adjoining locations, and in a more constraining way than merely being, like retinal ganglion cells, inhibited by neighbors without relinquishing their singular local signature. The study of how the several attributes of orientation, eye of origin, perhaps even color and disparity, can be accommodated anatomically while retaining some semblance of a retinotopic map, has raised awareness of the problem (Das & Gilbert, 1997). But once it is accepted that separate representations of these attributes are overlaid, resynthesis of a unitary point-for-point map of the original object space is no longer an adequate model. Using the sensation of a moving stimulus as an example: To register movement, it is necessary, of course, to have detected an association of spatial and temporal difference in stimulation, but once that has been achieved, a single neural token, now stripped of its generating temporal/spatial input patterning, suffices to store and signify the primitive element of movement, and even to have its magnitude and direction (though perhaps not its accurate location) finely discriminated.

### 6.2. Superposed apparatus for refining spatial signals

Performance in spatial tasks is much better than the rather coarse neural representation would imply. For example, even the most finely-tuned orientation-selective neurons respond over tens of degrees of orientation, yet orientation discrimination of lines is routinely measured to be just a fraction of a degree. The same applies to the relative location and disparity parameters of a stimulus where discrimination thresholds are many times better than the spacing and receptive field size of the mosaic of cells in the retina and cortex. Hence there exists a processing apparatus that refines signals and creates a virtual representation in the domains of individual attributes, apart from the retinotopic lay-out identified by anatomy and by receptive field location (Westheimer, 2008).

It is not necessary, however, to postulate a manifold of, say, contour orientations with the grain of the smallest discriminable steps. More appropriate it is to think of neural circuits operating on the signals of individual neurons; sometimes the concept of a vector sum is invoked (Kapadia, Westheimer, & Gilbert, 2000) wherein the state of the circuit, which can have quite a precise value, is developed as a result of a variety of interacting signals between the neurons composing it as well as from other neural centers. Thus the orientation attributed to a neural elaboration of a line stimulus will be a function of the contribution of the neurons directely affected by the stimulus, as well as of possible interactions from neurons elsewhere affected by neighboring stimuli (for example, due to lateral inhibition). In effect, though the measure associated with a particular attribute, e.g. line orientation, may be precise, it can at the same time also fail to be veridical.

It may actually be unnecessary that there is a separate apparatus that allows exquisite discrimination of the fine structure of each the pattern elements––line position, length and orientation, magnitude of angles––individually. Depending on how neuron-intensive they are, these circuits may be found to share resources rather than being implemented separately and independently in parallel. There may be just a few in each cortical location, available to and operating in the domain of the attribute for which precision is demanded at the moment. Switching would be provided by top-down tuning for this purpose. Elevation of thresholds in situations when more than one discrimination is attempted (Jiang & Levi, 1991) can be seen as evidence for such a view.

### *6.3. Assembly of patterns and union of signals from modules elaborating more than one individual primitive*

Location of points, length, separation and orientation of lines, magnitude of angles, all these have been shown to be perceived with great precision. Even though there may be cortical neurons selective for some, perhaps all of these attributes, the measure assigned to them is, as we have just argued, derived from an ensemble of neurons. A square, for example, has location of its four corners, length and orientation of its sides, four right angles, parallelism of its opposite sides. The percept of a perfect square thus embodies elements and requires coordination of (and prevention of dissonance between) signals emerging from the processing apparatus dedicated to each of these attributes.

### *6.4. Motion has to be abstracted*

The invariance of object location in the visual world with changes in gaze when the pattern representation has been shifted to entirely different cortical regions, may perhaps be separated from the present consideration, and so does processing of movement *per se*. But that spatial patterns and their component primitives are robust to overall motion while the eyes stay still during intersaccadic intervals requires additional properties.

If there is no need for a search within the nervous system for an internal reconstitution of spatial signals from the retina that replicates in all essential details our finely textured, well furnished and remarkably veridical visual space because, starting with a point-for-point mapping on the retina, the elements (primitives) of the geometrical structure of an object are disassembled and each subjected in parallel to its individual processing, how then is the unity of the structure preserved?

Consider the situation when a single visual stimulus is presented to the eyes, say a red square in a certain location in the frontal plane. After passing through the optical and retinal stages and transmission of action potentials into the cortex, and after processing in their respective refining circuits, the attributes of color, orientation, location, etc. will be represented by specific states in these circuits, distributed presumably in different cortical sites. For the percept to be that of a square and not a rectangle or rhomboid––and one's discrimination ability here is very good––the resulting signal for each of these constituent primitives has to have undergone a refining operation: the angles are right angles, the lines are parallel, horizontal and vertical distances are equal. The associated percept is a reflection of this array of states. The word "binding" is sometimes used to suggest that concatenation may actually occur by way of intra-cortical neural activity. Though it may be easier to envisage binding of signal identifying the states of the brightness and movement circuits with those for location and orientation, nevertheless it has to be accepted that binding is a necessary concept even within the array of spatial attributes because their differing and disparate extension cannot be simply accommodated in a single two-dimensional manifold.

Now, except for retinal rivalry, gross incompatibility does not occur in spatial perceptions. (Even in the Penrose and Escher illusions, there is no dissonance at the level of individual pattern segments; a figure's impossibility appears not pre-attentively but only on detailed contemplation of the whole.) It follows that an elaborate neural apparatus must exist to take care of possible discrepancies in melding the output of the processing the individual primitive spatial attributes. As just mentioned, the perception of a square requires an appropriate constraining of its constituents when that is necessary for coordination and resolution of possible dissonances within the ensemble of the states of the component circuits.

It is here that the conceptual connection with geometrical–optical illusion enters. First, because the signal about the magnitude of an attribute emerges from the operation of a circuit encompassing interaction both within and between ensembles of neurons receiving cognate stimuli, these signals can deviate from those in object space by virtue of the rules of these interactions (e.g. lateral inhibition). Second, when there is conflict between the imperatives of the various constituent primitives of a configuration, its resolution in the interest of perceptual unity of a geometrical structure requires that one constituent primitive yield priority to one of the others. Such a view of the origin of geometrical–optical illusions is quite different from the errors of judgment put forward by most of their discoverers and sponsored by Helmholtz and his adherents; it sees in them the rules of operation not of the mind but of neural processing at a relatively early stage in the brain.

## 7. Deconstruction of Poggendorff illusion

Deconstruction of the Poggendorff illusion can serve as a convenient illustrative example. Participant components include a transverse line *ABDF* (Fig. 7) whose continuity is interrupted by a set of inducing contours, here a pair of parallel lines *CG* and *HE*. The illusion manifests itself in the apparent lack of alignment in the percept of the two segments *AB* and *DF* when they are in fact perfectly aligned in the stimulus. Right from the beginning of the geometrical–optical illusions it has been contended that acute angles are seen as larger than they are. Applied to the current configuration this would mean that the perceptual counterpart of angles *ABC* and *EDF* are expanded in comparison to the real objects that generate them and that it is the orientation of line *AB* (and of line *DF*) that yields. It follows that the extension of line *AB* across the inducers would then intercept line *HE* at some point $D'$ which is higher than *D*. The interaction in the angle domain, enhanced magnitude, has to be postulated as affecting only the orientation of line *AB*, the other wing of the angle *ABC* being part of an extended and prominent line whose orientation is well anchored. What yields is the perceptual alignment of point *D* with line *AB*.

To make the sequence of the steps in this argument explicit:

(i) In the transformation from object space to visual space, angles *ABC* and *EDF* are enlarged.

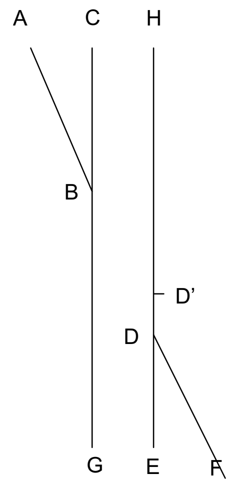


**Fig. 7.** The Poggendorff illusion is due to the perceptual expansion of acute angles. Angles ABC and EDF are perceived as larger than they are and consequently an extension DF of the transverse line AB, which actually intersect line HE at D, would need to be shifted to a higher position, say D', to be seen as aligned. This is an example of a geometrical–optical illusion interpreted as a manifestation of the resolution of a conflict, here between the relative effectiveness of orientation contrast for real (AB, DF) and for virtual (BD) lines.

(ii) Because line *CG* being more extended and more solidly anchored by virtue of its being yoked to line *HE* though the property of parallelism, it is the orientations of lines *AB* and *DF* that experience a shift in the transformation to visual space.
(iii) Such shifted orientations will produce the appearance of a lack of alignment of the two lines.

Supporting and confirming this line of reasoning, there are the findings that the magnitude of the Poggendorff illusion suffers the more elements of the configuration are omitted or weakened (Weintraub, Krantz, & Olson, 1980) or the more the saliency of contours is reduced (Tibber, Melmoth, & Morgan, 2008; Westheimer & Wehrhahn, 1997).

One of the observations on the Poggendorff illusion that stubbornly resists conventional interpretations can here be folded in: the illusion disappears when the pattern in Fig. 7 is rotated so as to make the transverse line either vertical or horizontal and the parallels oblique. If it is held that acute angles are perceptually enlarged, presumably arising from a repulsion in neighboring signals in the domain of the lines' orientation, this does not yet say anything, however, about how the supposed orientation shift is distributed between the two lines creating the angle: either, or both, or perhaps even just their virtual bisector? (That the virtual bisector plays a significant role, not always acknowledged, is demonstrated that it, rather than the component lines forming the angle, is the subject of the oblique effect (Li & Westheimer, 1997; Westheimer, 2003).) In the framework presented here, absence of the Poggendorff illusion for horizontal and vertical transverse lines must be interpreted as implying that perceptual enlargement of acute angles is now at the expense of the parallels which may indeed have yielded now that they are oblique. But they are not involved in the illusion which as a consequence is absent. When the transverse lines are in cardinal orientations it is their orientation that resists shift.

## 8. Geometrical–optical illusions as compromise, revealing hierarchical order

Such an understanding of the Poggendorff illusion illustrates how geometrical–optical illusions can be used as windows into early visual processing:

(i) the perceptual correlate of each primitive emerges from the operation of a neural circuit which operates on the incoming signals and within which there are interactions, such as pooling and repulsion of neighbors through lateral inhibition;
(ii) as the percept of larger patterns assembles itself, the states of these several circuits need to be interdigitated, and possible conflicts resolved, in the interest of unity in the perceptual construct.

When conflict resolution leads to a detectable shift in the signal emerging from one or more of the circuits, this reveals a hierarchical order, some holding firm and others yielding. A geometrical–optical illusion results when there is an overt effect. It need not be thought that the hierarchical order is rigid; interactions within processing modules, and between them, are fluid, and in any case they are all susceptible to top-down influences, which may modulate their relative strength according to wider attentional and cognitive imperatives.

Nonetheless, the ubiquity and universality of geometrical–optical illusions and their robustness to a wide range of visual situations allow some fairly general statements about hierarchical ordering. The strength of cardinal over oblique orientations and of explicit over implicitly-drawn contours has already been mentioned. The property of straightness does not seem overly dominant––surprisingly in view of its apparently fundamental nature, defying proof and hence at the heart of the axioms of geometry. In the Hering illusion (Fig. 8, top) there is a succession of intersections each mandating acute-angle expansion of different magnitude. Instead of engendering a segmentation of the contour, they are accommodated by preserving the contour's continuity and sacrificing its rectilinearity. Oppel's observation of induced curvature in a straight line by an adjoining curve (Fig. 8 bottom) fits into the same mold, except that here there is repulsive interaction between the orientation of line elements when they are just neighbors and do not intersect.

Further illustrations of the use of geometrical–optical illusions to guage the hierarchical order of primitives as they interact in

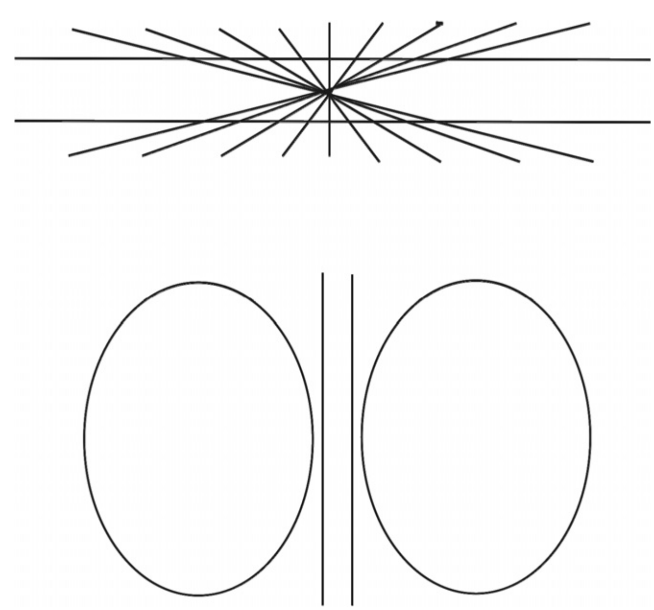

**Fig. 8.** Geometrical–optical illusions showing that the appearance of straightness and parallelism of a line pair is sacrificed to the demands of acute-angle expansion (top, Hering, 1861) or orientation–repulsion of adjoining contours (bottom, Oppel, 1854).

the generation of a unitary structure in the percept of a pattern are afforded by the Ponzo (Fig. 9) illusion, where length and separation of elements are dominated by oriented contours and not the reverse. A square ends up looking like a rhomboid, i.e., rectangularity and parallelism yield to the dictates of tilted contours. This takes place even when the contour's orientation is merely sketched in (Fig. 9 right).

Illusions caused by lateral inhibition in the domain of contour orientation, manifested principally by acute-angle expansion, are at the basis of a whole host of effects, of which the tilt illusion and the Zöllner effect are the most well-known. They permit quantitative studies relating to such contour variables as length, orientation, numerosity, contrast, chromaticity. The power of a geometrical–optical illusion in charting neural circuits is well illustrated by the simple example shown in Fig. 10. Explicitly-drawn lines are subject to the Zöllner illusion whereas point-locations are not. The two are inextricably linked in the geometry taught by Euclid yet neurally they are not processed in the same manner (Westheimer, 1996).

Because the spatial deformations of the geometrical–optical illusions depend on the establishment of signals for the primitives from which the patterns are constructed, it follows that these illusions would be affected when contour signals are impaired, as by diminishing contrast, or by restricting differences to only chromatic and not the luminance dimension. Geometrical–optical illusions generally persist under isoluminant heterochromatic conditions (Hamburger, Hansen, & Gegenfurtner, 2007; Li & Guo, 1995; Livingstone & Hubel, 1987), which as a topic should be seen as relevant not so much to the contours' geometrical properties as to the strength of their signals and to illusory effects in the brightness domain (e.g., Fig. 8).

When they are finally formulated, the rules governing the relationship between object-sided and perceptual geometrical properties of simple structures should begin to approach the goal of the Gestalt movement. Because they deal with primitive spatial elements and can be given quantitative expression by rigorous psychophysical experiments they can in turn act as guides in the analysis of the involved neural circuits.

Delineation (definition is too strong a word) of the primitives of spatial vision, i.e., their categorization, has had its origin primarily in perceptual judgments as sharpened by psychophysical experiments. Occasionally a serendipitous finding from neurophysiology has played a role in highlighting a perceptual category, for example the selectivity for contour orientation and movement in cortical neurons (Hubel & Wiesel, 1959, 1962). Yet the primacy of contours (Blachowski, 1913) and of movement (Wertheimer, 1912) in sensation had been stressed much earlier. On the whole, it is our subjective impressions on which we base decision of what in our perceptions is primitive, elemental, fundamental, and what is composite. Application of such knowledge to neurophysiological probing can then follow. In this geometrical–optical illusions will surely play a pivotal role.

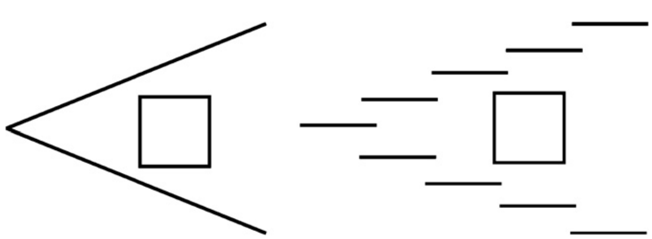

**Fig. 9.** A version of the Ponzo illusion which demonstrates how the relative orientation of the chevron's lines dominate and induce changes in the length of vertical lines, parallelism of horizontal lines and right angles to distort the appearance of a square object. It is surprising to recognize that parallelism and rectangularity yield to the dictates of contour orientation even when, on the right, this is only implicit.

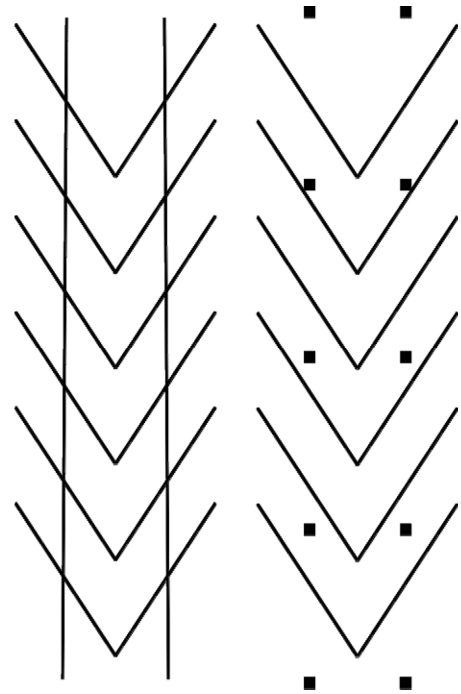

**Fig. 10.** Left: segment of the Zöllner illusion, showing how the perceived orientation and parallelism of the verticals is subsidiary to the mandates of acute-angle expansion. Right: illusion disappears when the vertical contours are no longer explicit but merely sketched in by point-pairs with equal separation.

## Appendix I. The beginning of geometrical–optical illusions, 1852–1863

Geometrical–optical illusions as a topic of its own arrived relatively late on the scene, described within a span of about ten years in the 1850's by several German physicists and physiologists each at an early career stage. There is every reason to believe that they all made their discovery independently. Johann Joseph Oppel (1815–1894), a Frankfurt high school physics teacher who also gained fame for his study of south-German dialects, is widely regarded as the founder of the subject. Indeed, in his seminal paper which appeared in the obscure annual report of the Frankfurt physics club (Oppel, 1854), he not only coined the word and described many of the phenomena but also clearly demarcated them from other visual illusions. They are not due to irradiation, because they show up equally in black on white as in white on black. Nor are they akin to the moon illusion which depends on the recognition of the relationship between the size of a familiar object and its distance from the observer. In a short contribution, Oppel describes vertical/horizontal anisotropy, the expansion of acute angles, filled intervals appearing wider than empty ones, straight contours appearing curved next to curved ones. He made what he called "Gegenversuche" which are attempts to null out an illusion and tried a quantitative approach by measuring relative dimensions in the drawings of pupils in geometry classes and demonstrating that the effects were more or less universal. He even looked at architectural features and interrogated artists and painters about their intuitive knowledge of the phenomena.

As original as Oppel's contribution no doubt was, his was not the first description of what he termed geometrical–optical illusions. An early and quite explicit mention of the vertical/horizontal anisotropy has been traced by Pastore (1971, p. 383)to a passage in

Nicolas Malebranche (1638–1715): "If a line is drawn on paper and another is drawn at its end perpendicular and equal to it, they will appear roughly equal. But if the perpendicular is drawn at its middle, the perpendicular will appear perceptibly longer, and the closer to the middle it is drawn, the longer it will appear (Malebranche, 1997)." In one of the more curious turns in the history of the multiple independent discoveries of the geometrical–optical illusions, Malebranche's observation seems to have been completely overlooked and remained unmentioned until Pastore pointed to it.

The role of another pioneer of this topic has also been universally ignored. Rudolph Hermann Lotze (1817–1881) in his influential book *Medicinische Psychologie* first published in 1852 (Lotze, 1852), specifically described the vertical/horizontal anisotropy and also that filled intervals are seen wider than empty ones. Because the book was widely regarded as important, it is surprising that it was never quoted by the others in their first description of the illusions. One of these was Adolf Fick (1829–1901) who devoted his MD Thesis at Marburg University of 1851, published in 1852 (Fick, 1852), to showing that a square object appeared oblong and that optical factors in the eye could not account for the differences. Fick right away eliminated some classes of explanation by recognizing that the illusions did not depend on observation distance and that they were present for white as well as for black patterns.

Within a very few years, the scene shifted to Leipzig, one of the foremost centers of scholarship at the time. It was Fechner's home, Oppel's alma mater and at one time or another housed the three other scholars who independently published ground-breaking papers in the subject between 1860 and 1863. These were Friedrich Karl Zöllner (1834–1882) at the time a physics post-doctoral in Berlin, Ewald Hering (1834–1918) a junior lecturer in physiology and August Kundt (1839–1894) a graduate student in physics in Leipzig.

Zöllner's illusion had its origin in a cloth pattern that he saw in his father's factory. When Poggendorff, editor of the most prestigious physics journal, read the paper he noticed also that the halves of the thick transverse cross hatches seemed spatially offset and mentioned this to the author who included this observation and gratefully acknowledged this help. Thus the Zöllner and Poggendorff illusions entered the literature simultaneously in 1860 (Zöllner, 1860).

A year later, in a monograph published in several sections and containing many of his seminal findings in spatial and stereoscopic vision, Hering (1861) also included a short section on geometrical–optical illusions, especially the wider extent of filled as compared with empty distances and the observation of curvature and non-parallelism of a line pair when embedded in a sheaf of radiating lines. Like Oppel and Zöllner before him, Hering looked to the apparent expansion of acute angles for an explanation. And two years later, Kundt published a paper full of many detailed observations and measurements on apparent length of lines (Kundt, 1863).

It surely was not coincidental that those participating in parallel in the discoveries––Lotze, Fick, Oppel, Zöllner, Hering and Kundt, to all appearances working independently––were at the time in their twenties and thirties, all born and educated in one of the central German states, and none had at the time an established academic position. It can be taken for granted that they were familiar with Kant's teachings and with Goethe's writings on color perception which popularized insight into perceptual experiences combined with a modicum of experimentation. The observations underlying geometrical–optical illusions can be seen as embedded in the dialectic which pitted these trends against the growth of objective, scientific––even mathematical––lay-out of natural phenomena in the physical and physiological arenas based on the concept of the existence of a real world with measurable content. Such roots of the discovery of geometrical–optical illusions in the mid-European, mid-nineteenth century culture have yet to be traced. It remains to be explained, for example, why Helmholtz, so wide-ranging in his contributions to physiological optics, played only a passive, reporting role, though he hastened to include this material in the 3rd part of his handbook (Helmholtz, 1867).

Geometrical–optical illusions quickly took off as an enterprise in visual science, yet it was decades before such names as Munsterberg, Ponzo and Muller-Lyer became part of the canon. References to many of the original papers are given in Hofman, (1920).

The compelling observations that these illusions were robust to observation distance and contrast polarity made all participants, from Fick on, reject optical factors in the eye. The majority, in particular Oppel writing well before Helmholtz popularized the concept, were content to posit errors of judgment as the cause of their effects. The exception was Hering, who in 1861, a decade or more ahead of expressions of similar views in connection with the light and color sense, put forward the proposition that the geometrical–optical illusions were expressions of misrouting of signals in the path through their neural processing, presciently pointing the way to the modern mode of their analysis.

## Appendix II. Geometrical–optical illusions viewed as arising from the transformation from object to visual space and their metrical properties

As an example of the application of Klein's program of laying out geometry as a study of structures and the transformations under which their properties remain invariant, consider at the outset the simple configuration of an equilateral triangle in a two-dimensional surface. In a Euclidean plane, its properties include that its sides have equal length and that it has equal interior angles adding up to two right angles, properties that remain invariant with what Klein calls the principal group of transformations: translations, rotations, zooming and mirror reflection.

Suppose now that this triangle is transferred to the surface of a sphere, also two-dimensional. Here, first of all, the concept of a straight line must be clarified to mean a geodesic, i.e., the shortest distance between two points, which now is a great circle. As regards transformations that allow the properties of the configuration to remain invariant, these include translations and rotations: a triangle made up of great circles of equal side length can assuredly be translated and rotated within the surface. However, the property of all angles adding up to 180° is now lost, and when such a triangle is enlarged by increasing the separation of its vertices, the interior angles, albeit equal, become still larger. The situation becomes more complex when the surface to which the Euclidean triangle is transferred is not that of a sphere, but of a football. Now the triangle, if it retains sides of equal length, may not always have all of its interior angles equal and moreover, these angles will change not only with zooming, as on the surface of a sphere, but also with most displacements within the football's surface.

Hence delicate issues arise about congruence of configurations as they are transferred from one space to another. A common way of looking at the situation, that of intrinsic curvature of a surface, is quickly understood in the two-dimensional case. A plane has zero curvature and the surface of a sphere's is constant, whereas that of a football varies from place to place and even from direction to direction in a single location. There is a close connection between the curvature and the distortions which configurations undergo on transfer between spaces. If the concept of intrinsic curvature of a surface is now being introduced in connection with apparent change of configurations and visual illusions, it must be understood that in differential geometry as here applied it is just a way of dealing with metrical properties within the surface;

it is a metaphor rather than an extension into a real third dimension. The analogy is apt with an ant crawling along the surface of a football and finding that the Pythagorean relation does not hold––tracing out the hypotenuse of a right-angled triangle may measure off a distance other than the square root of the squares of the other two sides, and the difference can vary from place to place. In fact, a measure of this difference at all points on the surface will give full characterization of the geometry of the surface without ever leaving it or even being aware of a third dimension.

This is precisely the attempt made here: a configuration in the physical fronto-parallel plane when it is transferred into visual space (i.e., when an observer describes the way it is seen) may undergo a transformation that makes it seen with an illusion. Is there a way of formulating any of the changed geometrical properties of the seen configuration in terms of the mathematical expression of the metaphorical "curvature" of the fronto-parallel surface? If there is, a full and satisfying phenomenal account will have been arrived at of such an illusion in terms of differential geometry.

The most general formulation of metrics of spaces is that of Riemann, who explained that the Pythagorean distance relationship $d^2 = \delta x^2 + \delta y^2$ in a Euclidean surface must be widened, for a more general two-dimensional case, to the form

$$d^2 = g_{11} * \delta x^2 + 2 * g_{12}\delta x * \delta y + g_{22} * \delta y^2 \tag{1}$$

where the $g$'s are parameters depending on the location and the coordinate system. In some special cases a coordinate system can be found for which they are invariant with location, and then the surface has the property that a fixed configuration can be freely moved on the surface without distortion. Examples are the Euclidean plane, of course (where in a Cartesian coordinate system $g_{11}$ and $g_{22}$ are equal to one and $g_{12}$ zero), and the surface of the sphere. In the more general case, when the $g$'s vary from place to place, one can think of the surface as being warped and then the distance between two locations separated by equivalent values of their coordinates differs over the surface.

The kind of approach had been introduced into the psychometric analysis of color space which provides a helpful illustrative example of what will be considered below for the space of object locations.

When the three-dimensional color space is constructed to make equiluminant stimuli a two-dimensional plane, the just-discriminable differences (j.n.d.'s) vary depending on the color location. To a first approximation they may be fitted by ellipses (Fig. A2.1) whose size and orientation can be expressed in terms of $g$ values in Eq. (1). These ellipses portray stimulus–space distances that are equal in the perceptual color space, for by definition all chromaticity j.n.d.'s are equal perceptual steps. Thus the ellipses in the equiluminant plane of physical color stimulus space can be transposed into circles on the equiluminant surface in perceptual color space by giving each element of the latter the tilt necessary to project the ellipse into a circle, and then piecing the elements together into a continuous surface. As now depicted (Fig. A2. 2), the latter has intrinsic curvature in the sense of differential geometry, but still remains a two-dimensional surface of equiluminant chromaticities. Stimulus coordinates, such as the spectrum locus can be marked on it. The virtue of this approach is that distances in the new surface relate directly to the perceived differences in chromaticity. For example, the equiluminant surface is constructed so that the distance traveling on the surface between 530 and 540 nm spectral loci is five times longer than between 600 and 610 nm, because psychophysically the former has about five times as many j.n.d. steps as the latter.

This approach to inventing a convoluted surface wherein intrinsic distances correspond to the perceptual distances between stimuli whose physical coordinates are laid out in a plane, was

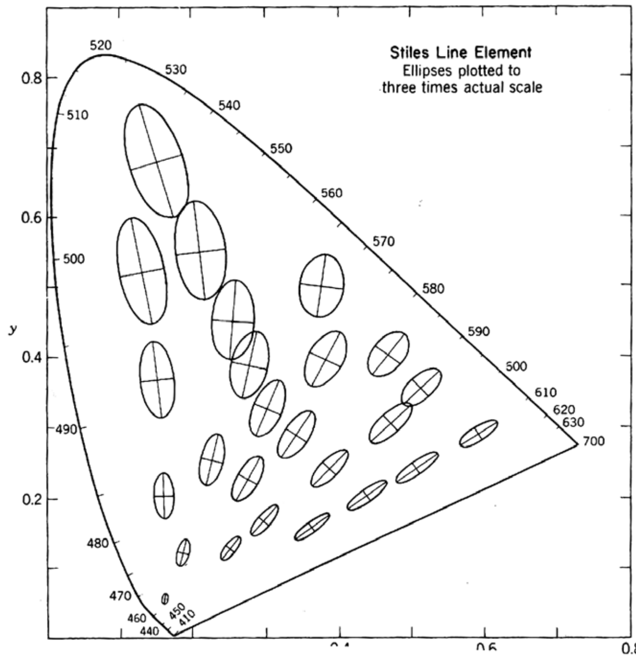


**Fig. A2.1.** Equiluminant plane in color space in which the just-detectable chromaticity differences have been plotted for various locations. To achieve a remapping in which the ellipses become circles of equal size requires the surface to be warped (see Fig. A2.2). From Wyszecki and Stiles (1982).

suggested originally by Helmholtz late in life (Helmholtz, 1896), when he tried to derive color differences from the Weber/Fechner law applied to the fundamental colors. It goes by the name of line element theory and is thoroughly explored in Wyscecki and Stiles's magisterial handbook (Wyszecki & Stiles, 1982). These authors make the distinction between inductive and empirical line elements. The properties of the curved surface derived from psychophysical measurements of chromaticity j.n.d.'s, though there is a certain amount of calculation involved in their generation, are empirical. The inductive procedure derives these quantities through a process that, also ultimately empirically-based, involves many assumptions that must always be stated explicitly to enable the validity of their utilization in the model to be evaluated. As compared to color space, where it has made considerable headway, this approach has not as yet been notably productive when applied to the visual space of object location.

An examination of the relationship between the physical object space and its representation in visual space can of course not be done without some markers. Translated to the representation in an observer's visual space of a simple fronto-parallel plane and phrased in the terms of Klein's Erlanger Programm, the inquiry concerns the extent to which geometrical properties of simple configurations of black markers on a white sheet of paper are preserved (remain invariant) in the observers' report of what is seen. The question "Are three actually collinear points reported as collinear?" is a specific instance of testing for invariance of the collinearity property during a transformation. On the whole, for small distances in isolation and fronto-parallel viewing the answer is that it holds. (That checkerboards covering visual angles of 100° of arc or more are seen as curved, a problem that exercised the minds of earlier generations of visual scientists (Tschermak, 1947), can be ignored in connection with the study of geometrical–optical illusions, where much more egregious distortions are evident in configurations extending just a degree of visual angle.)

The situation as regards the next example is also still promising. As it was known at least since Malebranche before 1700, equal

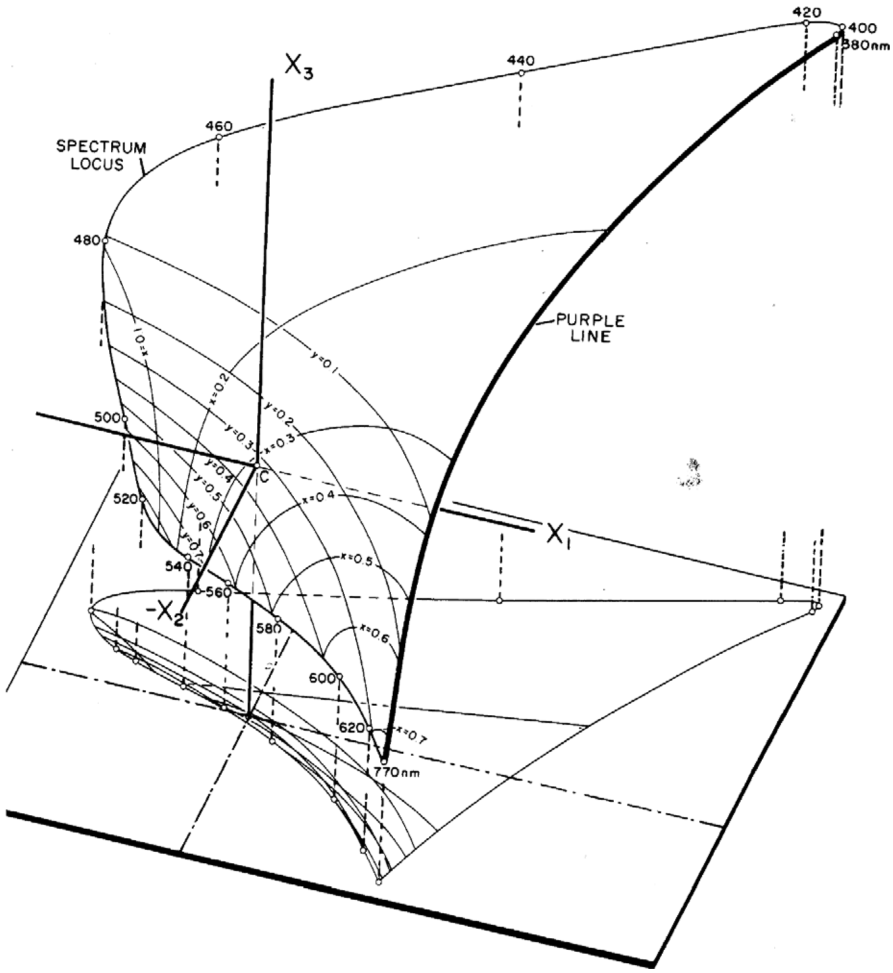


**Fig. A2.2.** Remapping of equiluminant plane in color space onto a surface in which just-detectable chromaticity changes have everywhere the same size. This representation is still two-dimensional, but has intrinsic curvature, i.e., is warped. From Wyszecki & Stiles (1982).

vertical and horizontal distances in physical space are not seen as equal: the vertical appears longer (Fig. A2.3). There are many instances of this illusion. In the same paper in which Oppel explicitly pointed to it (and in the process coined the term "geometrical–optical illusion") he offered a possible source: we routinely look down in front of us. By perspective, the vertical visual angle subtended by known figures when they lie on the ground before us is substantially less than in their direct view. Hence, in Oppel's words, "the eye" or, in Helmholtz's, "unconscious inference" routinely enlarges the actual vertical dimension of retinal images to compensate for this perspective shortening. It is an example of the kind of answer offered to the "Why" of visual illusions and distinguishes it from the physiological approach, the "How?" In this sections we are engaging in yet a different discourse, namely how the illusion might be viewed in the light of geometrical scholarship. As it happens, in this particular instance the answer is: quite easily. Just stretching one dimension by a constant factor in the transfer from object space to visual space still leaves it Euclidean, the surface in visual space representing the objective fronto-parallel plane is still a plane and straight lines remain straight though distances and angles may undergo changes. This kind of transformation is known as affine and the deformation suffered by configurations are least problematic of all that might be encountered; they can be accomplished by projecting markings located in the objective plane by parallel projection onto an imaginary tilted plane, rotated around a vertical axis (Fig. A2.4). Using this transformation, all real horizontal distances are seen foreshortened. It is understood, however, that this visual space equivalent of the real fronto-parallel plane will still appear fronto-parallel. The exercise of tilting is merely a metaphorical device to illustrate how the transformation––real object plane in which horizontal = vertical to visual fronto-parallel plane in which horizontal < vertical––might be brought about. Once physical horizontal dimensions of geometrical structures have been suitably adjusted, all operations can be carried out as before.

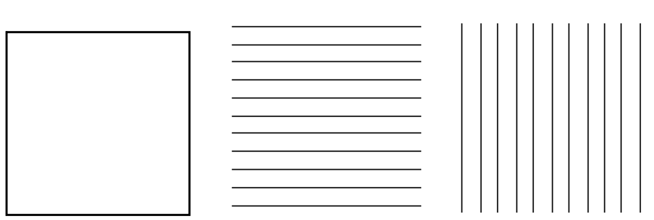

**Fig. A2.3.** A square looks taller than wider, and when filled with a stack of horizontal lines this is accentuated and when filled with a row of vertical lines it is reversed.

Plane tilt and parallel projection as the procedure for stretching one dimension with respect to an orthogonal one is a convenient way to envisage the transformation and serves well as a prototype of the wider type of operations. In general, the needed distortions require the projection onto a warped surface whose metric may change from place to place and whose measure is given by the local values of the intrinsic curvature. But both the original object plane,

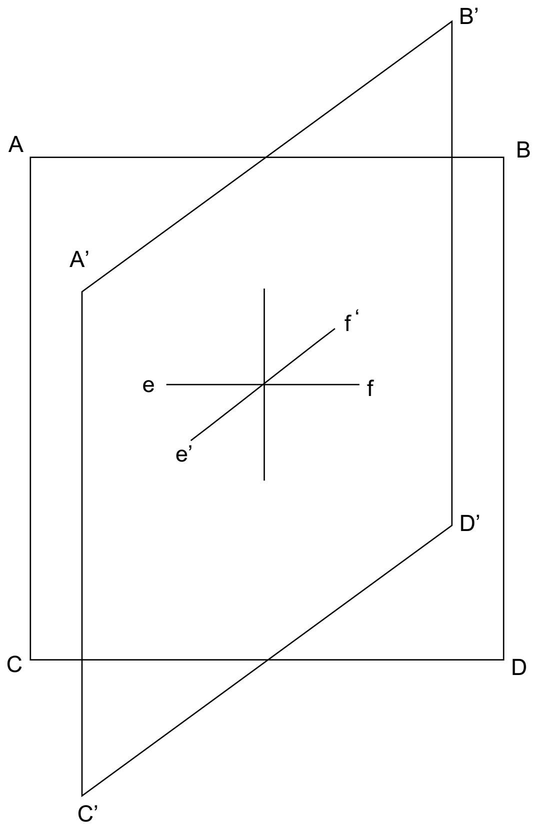


**Fig. A2.4.** Visualization of the transformation from object space to an observer's visual space that produces vertical/horizontal anisotropy. Frontal plane ABCD should be imagined as having been rotated around a vertical axis while still viewed head-on. The cross then will still be seen as right-angled, but its horizontal limb *ef*, now seen as *e'f'*, is shorter relative to the vertical.

and the perceived counterpart on which the geometrical structures have had their distortions generated in this fashion, are fronto-parallel in their respective spaces.

There is, however, a special difficulty attending this approach which is present neither in the above case, nor in that of the color space, nor even in the three-dimensional non-Euclidean visual space pioneered by Luneburg and his school (Luneburg, 1947). It relates to the fact that most geometrical–optical illusions are the result not of intrinsic distortions in the transformation from object to visual space, but depend on the content. Hence postulating merely a generalized stretching of vertical distances (or even an intrinsic non-Euclidean hyperbolic metric, as Luneburg did of three-dimensional visual space) does not suffice. A more relevant example is the disturbance of the metric of three-dimensional physical space due to mass in the theory of general relativity. Thus the simplicity and regularity demonstrated in this seminal exemplar of introducing a Riemannian metric in empirical science does not prevail here.

The nature of the problem is best exemplified by the geometrical–optical illusion mentioned by Lotze and described in detail by Oppel: filled spaces are seen larger than empty ones. If there were only one kind of filling effect, it might have been possible to postulate a warping of visual space that made distances longer. One can imagine a theory in which the presence of a discontinuity in the content of the space––say, a line or a border––becomes the equivalent of a fold in the surface. This is in principle a feasible mode of proceeding. In particular Helmholtz's pattern shown in Fig. A2. 3 would be amenable to this kind of calculation. But simplicity ended when Lotze pointed out that the phenomenon occurs over a wide range of filling patterns. Then the tasks would become one of relating their shape and contrast properties to the deformation they cause in the overall visual extent. Differential geometry can in principle deal with such situations by extending the formulations from cases in which the change in the intrinsic curvature of the fronto-parallel surface in visual space is continuous, to ones in which there are discontinuities, as indeed the change in refractive index (Snell's law) at a surface can be encompassed in the treatment of the trajectory of light rays as geodesics.

Change in apparent length are the defining feature of one of the most enduring of the geometrical–optical illusion, that known as the Mueller–Lyer illusion. The contextual components that can engender the illusion are extremely variegated and, moreover, as in many of these illusions, the distances need not even be laid down by explicit contours. Hence one can foresee that any investigation of this kind will soon become mired in detail and fail to provide the generalities and regularities expected when embarking on a geometrical mathematical exploration.

A very large class of geometrical–optical illusion is comprised of those in which straight lines look curved in the presence of neighboring contours or a sheaf of intersecting lines. In principle, this again can fall under the rubric of "warping" of the fronto-parallel surface in visual space: in transporting the straight line within an object-sided configuration into visual space, a hypothesized intrinsic curvature of the fronto-parallel surface in the latter would make a geodesic no longer appear straight. In popularizations of general relativity one sometimes sees the example of a heavy weight being tossed on a trampoline: straight grid lines on the trampoline surface in the unperturbed states now become curved.

This kind of analogy cannot be taken very far. Luneburg's enterprise, based on and supported by the relationship between the real and visual world locations of a few light points in the dark led to formulations which satisfied the sense of elegance, lucidity and concision of the mathematically inclined but failed to find traction once they encountered the richness, variety and complexity of everyday visual perception. The application of differential geometry to the study of geometrical–optical illusions seems destined to the same fate.

## References


Blachowski, S. (1913). Studien ueber den Binnenkontrast. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 47*(Abt. II), 291–330.

Coren, S., & Girgus, J. S. (1978). *Seeing is deceiving: The psychology of visual illusions*. Hillsdale, NJ: Erlbaum.

Das, A., & Gilbert, C. D. (1997). Distortions of visuotopic map match orientation singularities in primary visual cortex. *Nature, 387*, 594–598.

Dow, B. M., Snyder, A. Z., Vautin, R. G., & Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey. *Experimental Brain Research, 44*, 213–228.

Fick, A. (1852). Erörterung eines physiologisch-optischen Phänomens. *Zeitschrift für rationelle Medizin, 2*, 83–93.

Goethe, J. W. (1808). Zur Farbenlehre. Didaktischer Teil.

Hamburger, K., Hansen, T., & Gegenfurtner, K. G. (2007). Geometrical–optical illusions at isoluminance. *Vision Research, 47*, 3276–3285.

Hammer, E. R. (1949). Temporal factors in figural after-effects. *The American Journal of Psychology, 62*, 337–354.

Helmholtz, H. (1867). *Handbuch der Physiologischen Optik. Abt. 3*. Hamburg: Voss.

Helmholtz, H. v. (1896). *Handbuch der Physiologischen Optik*. Hamburg: Voss.

Hering, E. (1861). *Beiträge zur Physiologie I. Zur Lehre vom Ortssinne der Netzhaut*. Leipzig: Engelmann.

Hilbert, D., & Cohn-Vossen, S. (1932). *Anschauliche Geometrie*. Berlin: Springer.

Hofman, F. B. (1920). Die Lehre vom Raumsinn. In A. Graefe & T. Saemisch (Eds.). *Handbuch der gesamten Augenheilkunde* (3:1, pp. 1–213). Berlin: Julius Springer.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Jiang, B. C., & Levi, D. M. (1991). Spatial-interval discrimination in two dimensions. *Vision Research, 31*, 1931–1937.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology, 84*(4), 2048–2062.

Kitaoka, A. (1998). Apparent contraction of edge angles. *Perception, 27*, 1209–1219.

Köhler, W., & Wallach, H. (1944). Figural after-effects: An investigation of visual processes. *Proceedings of American Philosophical Society, 88*, 269–357.

Kundt, A. (1863). Untersuchungen über Augenmass und optishe Täuschungen. *Poggendorffs Annalen der Physik u. Chemie, 120*(30), 118–158.

Leibowitz, H. W., & Judisch, J. A. (1967). The relation between age and the Ponzo illusion. *American Journal of Psychology, 80*, 105–109.

Li, C.-Y., & Guo, K. (1995). Measurements of geometric illusions, illusory contours and stereo-depth at luminance and colour contrast. *Vision Research, 35*(12), 1713–1720.

Li, W., & Westheimer, G. (1997). Human discrimination of the implicit orientation of simple symmetrical patterns. *Vision Research, 37*(5), 565–572.

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement and depth. *Journal of Neuroscience, 7*, 3416–3468.

Lotze, R. H. (1852). *Medicinische Psychologie*. Leipzig: Weidemann.

Luneburg, R. K. (1947). *Mathematical analysis of binocular vision*. Princeton, NJ: Princeton University Press.

Malebranche, N. (1997). *The search after truth*. Cambridge: Cambridge University Press.

Morgan, M. J., Hole, E. J., & Glennerster, A. (1990). Biases and sensitivities in geometrical illusions. *Vision Research, 30*, 1793–1810.

Naka, K. I., & Rushton, W. A. H. (1966). S-potentials from colour units in the retina of the fish (Ciprinidae). *Journal of Physiology, 185*, 536–555.

Oppel, J. J. (1854–1855). Ueber geometrisch-optische Täuschungen. *Jahresbericht des physikalischen Vereins zu Frankfurt am Mains,* 39–47.

Pastore, N. (1971). *Selective history of theories of visual perception*. New York: Oxford University Press.

Penrose, L. S., & Penrose, R. (1958). Impossible objects: A special type of visual illusion. *British Journal of Psychology, 49*, 31–33.

Petry, S. J., & Meyer, G. E. (1987). *The perception of illusory contours*. New York: Springer-Verlag.

Pierce, A. H. (1901). *Studies in auditory and visual space perception*. London: Longmans Green.

Robinson, J. O. (1998). *The psychology of visual illusion Mineola*. NY: Dover.

Ross, H. E., & Plug, C. (2002). *The mystery of the Moon illusion: Exploring size perception*. Oxford: Oxford University Press.

Schwartz, E. L. (1977). Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biological Cybernetics, 25*, 181–194.

Tibber, M. S., Melmoth, D. R., & Morgan, M. J. (2008). Biases and sensitivities in the Poggendorff effect when driven by subjective contours. *Investigative Ophthalmology & Visual Science, 49*, 474–478.

Tschermak, A. (1947). *Einführung in die Physiologische Optik*. Wien: Springer-Verlag.

Valeton, J. M., & van Norren, D. (1983). Light adaptation of primate cones. *Vision Research, 23*, 1539–1547.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*, 1260–1262.

Wade, N. (1982). *The art and science of visual illusions*. London: Routledge.

Weintraub, D. J., Krantz, D. H., & Olson, T. (1980). The Poggendorff illusion: Consider all the angles. *Journal of Experimental Psychology. Human Perception and Performance, 6*, 718–725.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie, 61*, 161–265.

Westheimer, G. (1996). Location and line orientation as distinguishable primitives in spatial vision. *Proceedings of the Royal Society of London. Series B, Biological sciences, 263*(1369), 503–508.

Westheimer, G. (2003). Meridional anisotropy in visual processing: implications for the neural site of the oblique effect. *Vision Research, 43*(22), 2281–2289.

Westheimer, G. (2006). Specifying and controlling the optical image on the retina. *Progress in Retinal and Eye Research, 25*, 19–42.

Westheimer, G. (2007). Irradiation, border location and the shifted-chessboard pattern. *Perception, 36*, 483–494.

Westheimer, G. (2008). Hyperacuity. In L. A. Squire (Ed.), *Encyclopedia of neuroscience*. Oxford: Academic Press.

Westheimer, G., & Wehrhahn, C. (1997). Real and virtual borders in the Poggendorff illusion. *Perception, 26*(12), 1495–1501.

Wohlwill, J. F. (1960). Developmental studies of perception. *Psychological Bulletin, 57*, 249–288.

Wyszecki, G., & Stiles, W. S. (1982). *Color science*. New York: Wiley.

Zöllner, F. (1860). Ueber eine neue Art von Pseudoskopie und ihre Beziehung zu der von Plateau und Oppel beschriebenen Bewegungsphänomenen. *Poggendorffs Annalen der Physik u. Chemie, 110*(186), 500–523.