

Vision and Color

M. A. Webster

The mysteries of seeing have caught the eye of many renowned physicists. For example, Newton, Young, Maxwell, and Schrodinger were all central figures in the history of color science [20]. This chapter brings up to date a previous version written by the color scientist Edwin Land, who was himself renowned as an inventor and founder of the Polaroid Corporation, but who maintained an active interest in the psychology of human color vision throughout his career. Land developed a series of intriguing and controversial demonstrations of color perception [25]. In a typical example, grayscale images of a scene were photographed through either a red or a green filter. The two positive black and white slides were then recombined on a projection screen, with one projected through a red filter and the other with no filter. Though this restricted the stimulus to only “red” and “white” illumination, the resulting display took on a wide range of varying hues – and importantly – tended to reproduce the colors of the original scene. Newton’s experiments with prisms led him to conclude that “the rays are not colored” and thus that color is a property of the observer [35]; and many demonstrations like Land’s have shown moreover that the sensation of color does not correlate in a simple way with the spectrum of light at each point in the scene, for it also depends fundamentally on the surrounding context.

This chapter places such demonstrations within the context of general principles that are thought to have shaped the design of the *visual system*, and illustrates these with specific examples of the processes mediating *color vision*. For detailed reviews covering a wide range of topics in vision, *The Visual Neurosciences* edited by Chalupa and Werner [6] is highly recommended. Recent comprehensive discussions of color vision can also be found in Mausfeld and Heyer [31], and Gegenfurtner and Sharpe [10]. For basic data on color and its specification and measurement, *Color Science* by Wyzescski and Stiles [49] remains an invaluable resource, while more recent data are available through the Color and Vision database at www.crv1.org. Finally, the Viperlib Library at www.viperlib.org has become a central site for demonstrations and illustrations related to vision science.

Entry for list of contributors:
Michael A. Webster
Department of Psychology,
University of Nevada, Reno, NV,
USA
VISION AND COLOR

Revision date: 03-01-2005

Sensory Coding

Channels and receptive fields

At early stages of sensory systems, information about the stimulus is encoded and represented in terms of multiple “channels”. A channel is like a template or selective filter, that responds to a limited range of levels along some stimulus dimension [48]. For example, auditory channels encode different ranges of sound frequency, while visual channels are selective for properties like the size or orientation of edges. In this way information about the stimulus can be represented in terms of the distribution of activity across a bank of filters. It should be emphasized that a channel is a theoretical construct, typically defined by behavioral experiments. For example, if the ability to hear one tone is masked in the presence of another, then this is evidence that the two frequencies are detected by a common sensory channel [33]. Widely separated frequencies do not interfere, suggesting they are carried by different channels. Channels defined in this way are not equivalent to actual neurons, and how they are related to neural activity remains uncertain. Neural responses can be directly measured by recording the electrical

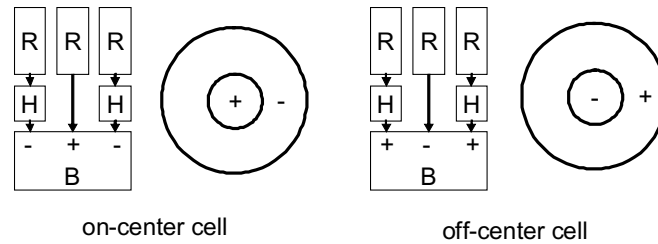


Fig. 1: A simplified schematic of lateral inhibition in the retina. Left: In an on-center bipolar cell (B), receptors (R) in the center of the receptive field directly excite the cell while receptors in the surround inhibit the cell via horizontal cells (H). The receptive field, or area over the visual field to which the cell responds, thus has a center-surround arrangement. As a result, an on-center bipolar cell is strongly excited by a bright spot of light (that fills the center) on a dark background, while strongly inhibited by a dark spot on a bright background. A large spot, that fills both the center and surround, elicits little response. Right: off-center cells instead have inhibition in the center and excitation in the surround, and thus are excited by a dark spot on a bright background.

activity of a cell while presenting different stimuli. Sensory cells also behave like selective filters, responding to some stimuli but not others, and these selective responses serve to define the “receptive field” properties of the cell [7].

For both channels and receptive fields, the selectivity of perceptual mechanisms tends to increase at successive stages of analysis [7]. Thus at the level of the photoreceptor where light is initially absorbed, the responses are selective only for the location of the stimulus (either light falls on the receptor or it does not) and the wavelength. Yet at the first synapse there is already considerable communication between receptors, forming receptive fields that can detect not just the presence of light but discontinuities in light, or contrast. Contrast coding is a consequence of “lateral inhibition” – cells receive excitatory inputs from some receptors while inhibitory inputs from others, and this spatial opponency is organized in a center-surround arrangement [24] (see Fig. 1). The effect of this receptive field structure is that the response now carries information about the net difference between the receptors, but at the cost of losing information about the absolute light level. For example, such cells might give a large response to a bright spot on a dark background (that just fills the excitatory center), while little response to a large bright field (since the opposing center and surround inputs then cancel). Similar cells are found with inhibitory centers, thus giving the largest excitatory response to dark spots (on a bright background). Thus in one step the visual system goes from a photodetector that registers light, to a pattern detector that encodes the variations in light across the image. At still higher levels the patterns required to stimulate a neuron become more complex. For example, in work that earned the Nobel Prize, Hubel and Weisel discovered that most cells in primary visual cortex will respond only to a light-dark edge that has a particular orientation [16].

How a channel or neuron conveys information is not definitively known. One obvious way is by the size of the response. For example, an edge at the preferred orientation may elicit strong activity, with responses declining as the tilt varies away from optimal.

The response magnitude may therefore provide a measure of the likelihood that the preferred stimulus is present, and a particular stimulus might then be represented by the mean or modal response across the set of channels. A second potential way is by the timing of the activity. The firing of auditory neurons can follow the oscillations of the sound stimulus at low and mid-range frequencies, and this phase-locking is thought to be important for pitch perception [33]. In the visual system temporal coding has recently received attention with regard to the synchrony in firing between different cells, and has been theorized to play a role in binding together the responses that encode different parts of the same object [9]. Finally, sensory cells must also carry information by the pattern of connections they form. That is, cells in the visual system are somehow “labeled” to convey visual information, for the same pattern of firing within a different sensory modality creates a completely different sensation. An example of this is visual phosphenes [5]. If you lightly tap the corner of your eye then a flash of light can be seen, for the visual receptors – no matter how they are stimulated – convey information about light.

Coding efficiency

If information is represented in channels, then this raises the question of which set of channels might be optimal for representing a given stimulus dimension. The answer to this is constrained by many factors. One is the importance of the stimulus to the observer. For example, for foraging primates ripe fruit may have special significance, and as discussed below this probably strongly influenced the evolution of primate color vision [38]. Similarly, many species have biologically important “trigger stimuli” for which special feature detectors have evolved. A still controversial example is the human face, which may be encoded by neural mechanisms that are specialized for discriminating the subtle physical differences that distinguish different faces or expressions [21].

Another important set of factors constraining the choice of channels derives from information theory and the assumption that visual system is trying to form a probabilistic model of the environment [43]. This focuses on how to make the most use out of the limited signaling capacity of neurons, which is restricted because neural responses are noisy and have a small dynamic range. One prediction from this approach is that sensitivity should be matched to the expected range of signals in the environment. If a neuron can reliably signal only a small set of levels, then those levels should be ones that discriminate between frequent levels in the stimulus, and not wasted to encode events that are very rare. This can be achieved by choosing the response function so that each level has an equal probability of occurring. Similarly, the probability of responding should also be equated across different mechanisms, again so that the channels each carry a comparable load. A further powerful prediction is that the responses of different channels should as far as possible be uncorrelated, so that each channel carries independent information. This is similar to the motivation driving image compression algorithms, which seek to represent the image with the least number of bits. Finally, a recent powerful approach has been to model the channels predicted by sparse coding. This tries to achieve a representation in which each channel has a low probability of responding, so that when a channel does respond it carries high significance for a particular image property. There are now numerous examples where these principles have been shown to closely approximate the channel structure observed at early levels of visual coding [36].

Parallel pathways

At a much coarser level, the visual system also has a channel structure insofar as information about different visual attributes is encoded by different processing streams [8]. These streams involve neural pathways that flow more or less in parallel through different stages of the visual system. One major pair of pathways is the parvocellular (p) and magnocellular (m) streams, which are named after the layers of the thalamic nuclei through which they project but which are formed as early as the retina (the neural layers lining the back of the eye which include the photoreceptors) [27]. Cells in the p pathway are sensitive to color and have high spatial resolution, suggesting they are the early building blocks of channels encoding color and fine-detail form perception, while m cells are relatively insensitive to color but can respond to very high temporal frequencies. The m cells may subserve a pathway specialized for encoding motion and visual space, though the actual functions and the degree of segregation remain controversial. A second general distinction is between the “what” and “where” pathways which stream through temporal and parietal cortex and reflect processes specialized for object perception and spatial relationships [45]. Evidence for these distinct pathways includes localized brain injuries which result in selective visual deficits, including the inability to recognize objects or faces, or to attend to different regions of space.

Numerous extrastriate visual areas have been identified beyond primary (striate) visual cortex, each with a more or less explicit map of the visual field but with very different response selectivities [8]. The number and nature of these areas remains debated, but it seems clear that the visual system includes a number of subsystems that are specialized for different types of analysis. This modular organization may be largely motivated by the fact that different sensory attributes require different neural computations and by the need to communicate between the cells involved in these computations. If a population of cells is required to solve a perceptual problem, then if nothing else, bringing them together within a visual area reduces the axonal mass required to connect them.

Color Vision

Cones

With these general concepts in mind we can now turn to the specific example of color coding. The human visual system is sensitive to a minute fraction of the electromagnetic spectrum, extending roughly from 400 to 700 nm. The choice of this range may have several causes. For example, the energy at shorter, ultraviolet wavelengths can damage the eye, and detecting them could degrade the image because the human lens does not correct for chromatic aberration due to wavelength-dependent refraction. We are largely insensitive to these wavelengths because they are filtered by a pigment in the lens. Sensitivity to longer, infrared wavelengths might cause the receptors to respond to the body’s own thermal energy, again degrading information about the outside world.

Within the visible range, the visual spectrum is sampled by three types of cone photoreceptors that have photopigments that absorb maximally at short (~ 440 nm), medium (~ 545) or long (~ 565) wavelengths. These S, M, and L cones thus form a three channel or trichromatic system for representing color. One of the most important insights into color vision is the “principle of univariance” which holds that any single cone type is color blind, in that it

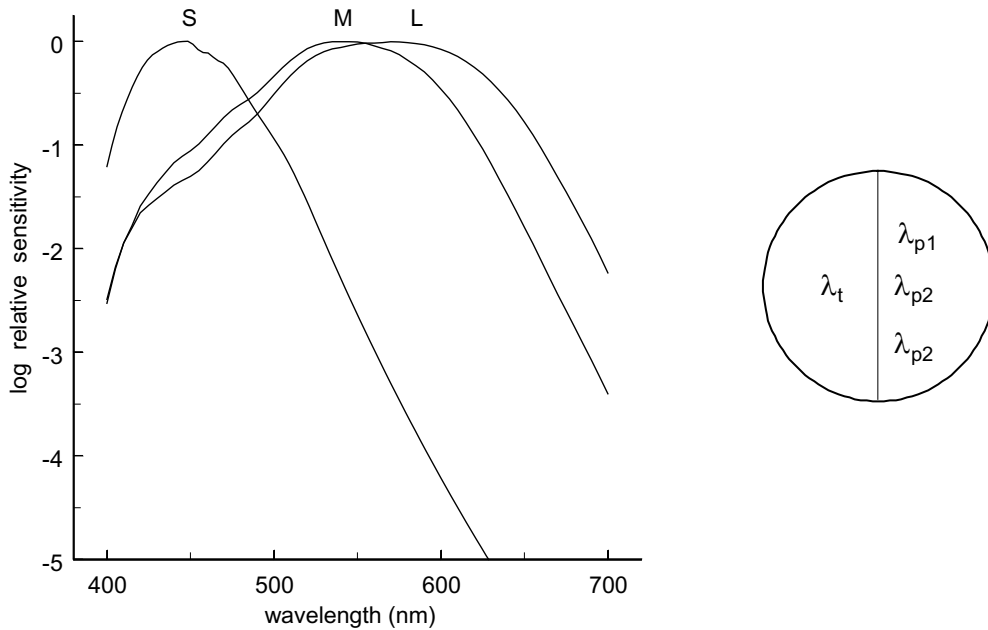


Fig. 2: Left: The spectral sensitivities of the short (S), medium (M), and long (L) wavelength sensitive cones. Right: Because the three cones provide only three univariant samples of the spectrum, any spectral light (λ_t) can be matched by an appropriate mixture of the intensities of three fixed-wavelength primaries ($\lambda_{p1} + \lambda_{p2} + \lambda_{p3}$). The match occurs when the lights on the two sides of the field produce equal absorptions in each cone class.

cannot discriminate a change in wavelength from a change in intensity [42]. This is because wavelength can affect the probability that a photon is absorbed by the photopigment, but once absorption occurs the isomerization or change in the molecular configuration is a discrete stereotyped event, and thus loses all information about the photon's wavelength. The receptor can therefore respond in only one way regardless of the wavelength. It is for this reason that we are completely color blind at very dim light levels, where only the rod photoreceptors are active. All rods contain the same photopigment, rhodopsin, which absorbs maximally at ~ 500 nm. Shorter or longer wavelengths produce weaker responses, yet if we increase the intensity of these wavelengths then we can find a level that exactly matches the response to a given intensity of the 500 nm light. Once the match occurs the two physically different lights produce exactly the same visual response and are thus indistinguishable, or metameric. By setting brightness matches between different wavelengths at very dim (scotopic) light levels, it is possible to accurately measure the spectral sensitivity of the rods.

While each cone type is similarly color blind, comparing the responses across different cone types allows the visual system to unconfound wavelength and light level. For example, consider two lights at 545 and 565 nm, near the peak wavelengths for the M and L cones (cf. Fig. 2). For either cone alone it is possible to adjust the relative intensities of the two lights so that the response to both wavelengths is the same. Yet no matter what the intensity, the

response to the 565 nm light will be larger in the L cones than M, and thus by comparing the L and M responses, the two wavelengths can be discriminated no matter what their intensities. On the other hand, the excitation of both L and M to 565 nm – or to any other wavelength or spectral distribution – can be matched by adjusting the intensities of two primaries instead of one. This is because the two univariant cones provide only two degrees of freedom, and the match thus reduces to two equations defining when for each cone the response to the test light and to the mixture of the two primaries is the same. This crude form of color vision based on two cone classes is known as dichromacy, and is a typical form of color blindness, in which one of the three cone photopigments (mostly L or M) is not expressed because of a genetic error [20]. Normal human vision is instead trichromatic, or three dimensional. The colors a trichromat can distinguish that a dichromat cannot are for spectra that are matched for two of the cone classes, but differ in their effects on the third cone type (the one the dichromat lacks). These are the color pairs used in standard color screening tests. With three cones the normal observer represents the spectrum with three degrees of freedom, yet note that this means that only three primaries are now required to match any spectrum. Thus a television can mimic the color of most spectra by varying the luminance of only three phosphors (and the ones it cannot are only because the saturation is outside the gamut of the monitor). Measures of color matching for spectral lights date back to Maxwell, and analyses of these matches still provide the most accurate estimates of the spectral sensitivities of the cones [44].

Trichromacy or even higher-dimensional color vision is common in many species, yet among mammals it is restricted only to a subset of primarily old-world primates [18]. Analyses of the similarities between different photopigment genes suggests that primate trichromacy may have arisen from a relatively recent evolutionary event (30–40 million years ago) leading to distinct L and M cones [34]. The L and M genes lie in tandem on the X chromosome and are 98% homologous, while the S gene is autosomal and only 43% similar to L or M. Thus like most mammals, our ancestors may have had only a single “ancient subsystem” for color based on comparing S cones and an ancestral L/M cone. The split of the latter allowed primates to evolve a second “modern subsystem” based on comparing the L vs. M cones. It turns out that this comparison is optimal for discriminating ripe fruit from foliage, suggesting that frugivory was a major driving force in our color vision [32]. In some species of new world primate there is only a single gene locus on the X chromosome but multiple alleles coding different pigments in the population. All males in these species are dichromats, yet females who inherit different alleles on their two X chromosomes express both and can be shown behaviorally to be trichromatic [18]. Thus evolving an extra color dimension may require only that extra information in the form of a new receptor type be made available. The rest of the brain may be plastic enough to automatically take advantage of this.

Color opponency

The signals from the three cone types are not kept segregated but instead are mixed within the retina to form channels that respond to different combinations of the cones. Such channels make sense because color vision requires comparing the responses across different cones. They have also been rationalized by noting that the responses of the cones themselves are highly redundant, because the cones have broad and closely overlapping sensitivities to wavelength and thus their responses to different spectra are very highly correlated. Postreceptoral mechanisms may recode the receptor responses into channels which are more nearly indepen-

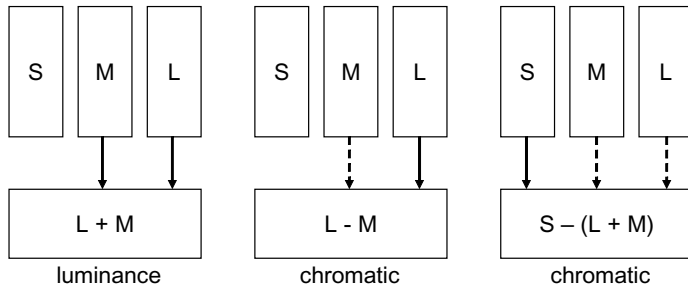


Fig. 3: A schematic of early stages of color coding in the human visual system. Signals from the L, M, and S cones are combined to form a nonopponent channel sensitive to luminance, and two opponent chromatic channels that respond to differences in the L vs. M cones or in the S cones opposed by the sum of L and M cone signals.

dent and which may allow the visual system to give more equal weight to brightness variations (which can be very large from one cone to the next) and chromatic variations (which result in much smaller cone contrasts again because of the overlapping spectral sensitivities) [29].

Early postreceptoral color channels fall into three classes [23], see Fig. 3. One adds signals from the L and M cones, forming a nonopponent channel responsive to luminance contrast (variations in intensity) but not chromatic contrast (variations in wavelength). S cones contribute very little to luminance, and in any case make up only about 5% of the total number of cones. S cones thus appear specialized for carrying chromatic information, while the L and M cones must be used for both color and luminance [26]. The remaining two channels receive antagonistic inputs from different cones, forming “color opponent” channels. One channel is excited by L cones and inhibited by M or vice versa, while the second is excited by S cones while inhibited by a combination of L and M cones. These channels thus give responses of opposite polarity to different parts of the spectrum. For example, a +L vs. $-M$ cell will be excited by wavelengths that produce a larger response in the L cones, while inhibited by lights that more strongly stimulate M cones. Moreover, a white light, that produces equivalent responses in L and M, will not excite the cell, since the opposing inputs will cancel. The cell thus codes which cone responds more while discounting the overall intensity. Note that the S vs. LM and L vs. M comparisons correspond to the ancient and modern subsystems predicted by the genetics of the photopigments. Note also that this color opponency is similar to the spatial opponency underlying the encoding of spatial information in center-surround receptive fields (cf. Fig. 1). In fact the different cone inputs may be segregated into the center and surround, so that the cell has both spatial and color opponency.

Perception

While the early analysis of sensory information is closely tied to the physical properties of the stimulus, these are often not the properties that we actually perceive. This is clearly illustrated in Land’s demonstrations of color vision. The colors observers report for an object cannot be predicted from the spectrum or from the cone absorptions at that point in the image, but do tend to preserve the color of the original photographed surface. This would be surprising

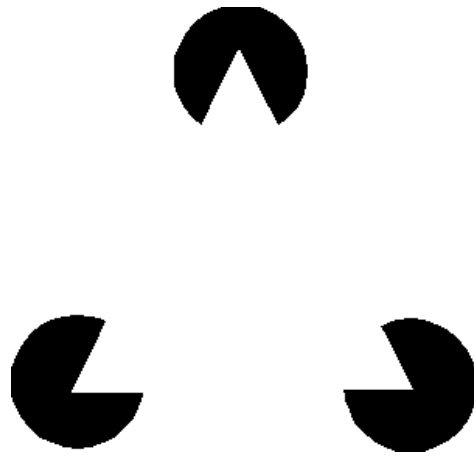


Fig. 4: Kanizsa's triangle. The three notched circles appear to be full circles occluded by an overlying triangle. The edges of the triangle between the circles are illusory contours.

if the purpose of the visual system was to measure light, but is less paradoxical when one considers that the function of vision is instead to tell us meaningful information about objects – the “what” and “where” of the stuff we are looking at. Light is the medium through which we visually acquire this knowledge, but the ecologically relevant information is not the energy reaching the eye (the proximal stimulus) but rather the surface that reflected or emitted it (the distal stimulus), and it is the latter to which perception more closely corresponds. In the case of color this means that color appearance tends to mirror the actual reflectance properties of objects rather than the raw spectrum of the light. That is, when we say that a banana is “yellow”, we are referring to a property of the banana and not of the light reaching the eye.

Perceptual inference

However, the general challenge faced in perception is how to estimate the object from the image. This is a problem because the image itself is inherently ambiguous, in the sense that many different stimuli could give rise to exactly the same pattern of light on the retina. For example, the retinal image is two-dimensional, and thus any given image could have arisen from the projection of an infinite number of different three-dimensional layouts. In the case of color, ambiguity arises because the spectrum reaching the eye is a product of both the surface reflectance and the illuminant, and thus the same distribution of wavelengths could be produced by different combinations of lights and objects. To resolve this uncertainty, the visual system has to make its “best guess” about the probable source of the stimulus, by using what Helmholtz termed “unconscious inferences” about the world [13]. A currently popular approach is to formalize the solution in terms of Bayesian inference [22]. While many distal stimuli are possible, they are not all equally likely, and these prior probabilities can be used to guide the selection among different possible interpretations.

Figure 4 illustrates the role of inferences in perception. The image consists of three circles with a sector removed, yet the perception is of three full circles with a triangle lying on top. Note that the visual system completes this figure both “amodally” (adding in the occluded

sections of the circles) and “modally” (adding illusory contours to complete the triangle). The resulting triangle appears as a solid figure that is brighter than the ground, indicating that it is not merely inferred but explicitly represented, and in fact it is possible to stimulate cortical cells with these illusory contours [37]. Similarly, completed occluded sections can be shown to influence visual performance, for example in a task where subjects have to search for a unique item, again suggesting that the visual system builds a representation even for the parts of objects that it cannot “see” [12]. This makes sense if the goal of perception is to guess the actual objects from the image. The probability that three independent objects would have precisely aligned sectors is low, while the probability that aligned edges are associated with a common object is high. (It might seem unlikely that an object would be missing its edges, but in actual images the object contours are often poorly defined, so that incomplete contours are common.)

Perceptual constancies

A problem closely related to perceptual inference is how to maintain a stable perception of objects despite changes in the images they project onto the retina. For example, the perceived size or shape of an object should remain invariant with changes in distance or viewing angle, yet these changes systematically alter the size and shape of the image. Size constancy is possible because there is a lawful relationship between visual angle and the size and distance of an object, and because the visual system is able to infer object distance in order to solve for the actual physical size. There are many different cues available for estimating depth, yet it is nonetheless striking how fully we perceive the three-dimensional layout of the world by reconstructing it from the flat images on our eyes [15]. It feels easy and intuitive to judge whether two objects are the same size when we view them at different distances, but this is only possible by discounting the differences in the visual angles they subtend. Conversely, it feels difficult and unnatural to choose a near and far object that subtend the same angle, even though the raw visual angles are directly available on the retina and must be factored in to any judgments about size.

Color Constancy

Lightness constancy. In color and lightness constancy, the problem is how to discount the illumination in the scene in order to infer the material properties of the surface. That is, the visual system has to separate out the contributions of the illuminant and the surface to the color signal reaching the eye (Fig. 5). Lightness refers to the extent to which a surface appears to reflect the ambient light. Consider the page before you. In indoor lighting the paper has a high lightness while the print appears dark. If we view the page in direct sunlight then the flux reaching the eye can be orders of magnitude higher, to the point that more light is now reflected from the print outdoors than from the paper indoors. Yet the print still appears black and the paper white, because in both cases the paper reflects much more of the incident light. Thus lightness has little to do with the absolute light level while much to do with the relative light levels in the scene, or the contrasts.

There are a number of mechanisms that contribute to lightness constancy. Two important ones are adaptation and induction. The sensitivity of the visual system is not fixed, but instead varies continuously to adjust for properties of the stimulus. These adjustments occur at all levels of the visual system and in response to many different attributes of the visual environ-

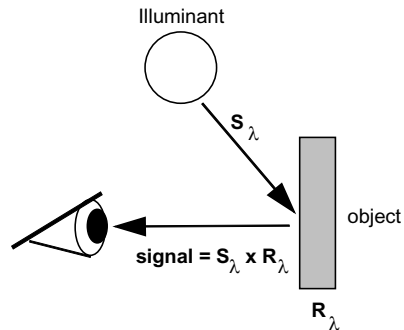


Fig. 5: The color signal reaching the eye from a diffuse reflecting surface is a product of the surface reflectance (R_λ) and the illuminant spectral power distribution (S_λ). Thus to estimate the surface color the visual system has to discount the effect of the illuminant on the cone absorptions.



Fig. 6: An example of light adaptation. Stare at the cross in the left figure for 30 seconds and then shift your gaze to the center of the gray square. An afterimage of the girl will be briefly visible and can be revived by blinking. The aftereffect arises because each local area of the retina adapts to the average light level it is exposed to, becoming more sensitive where the adapting image is dark and less sensitive where it is bright. The more sensitive areas thus respond more strongly to the gray square, resulting in a negative afterimage of the original picture.

ment [47]. For example, if you view the downward flow of a waterfall then the rocks to the side will appear to move upward, and this is because the visual system has rapidly adapted to the direction of motion in the scene. Light adaptation adjusts to changes in the mean light level in the scene, and may occur as early as the photoreceptors or at their synapses (Fig. 6). These adjustments are important because the range of lighting encountered far exceeds the range that could be coded by a fixed neural response, much as a camera's settings must be varied to bring the image within the available operating range of the film. However, the adaptation also

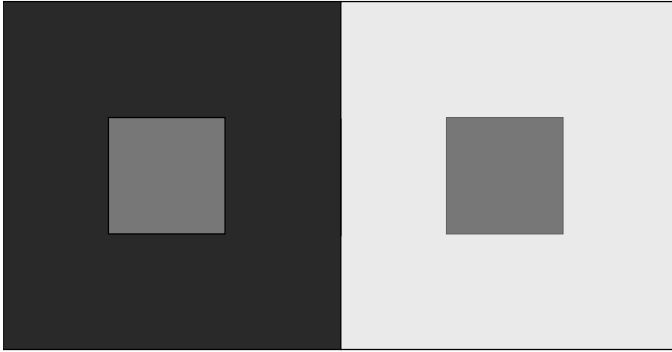


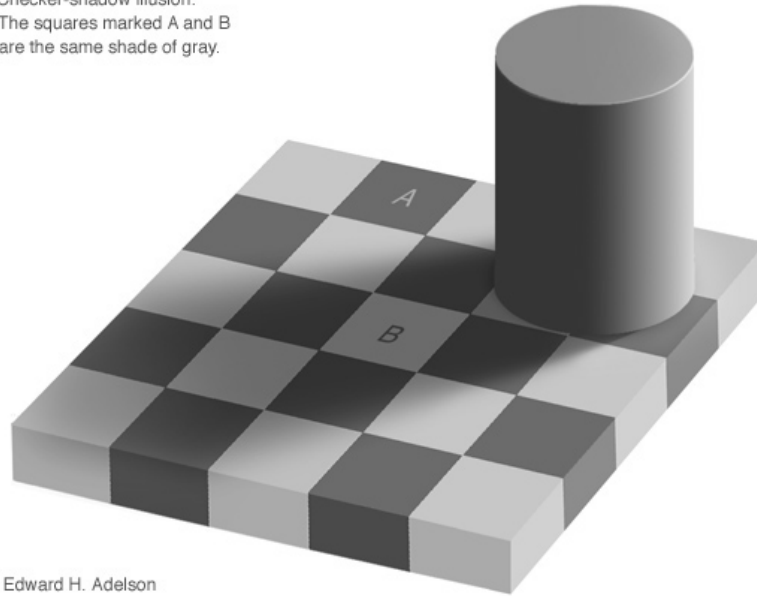
Fig. 7: An example of brightness induction. The two gray squares are physically the same, yet the square on the dark background appears brighter.

plays an important role in perceptual constancy. A simple form of light adaptation involves multiplicative gain changes that are reciprocally related to the average light level. That is, if the light is doubled then sensitivity is halved. In principle, a process of this kind would completely factor out a change in overall light level, leaving the responses to the contrasts within the scene intact. (Note that because the adaptation is spatially local, in order to adjust to the average light level the mechanisms must sample different parts of the image over time through eye movements. If one instead maintains fixation then afterimages result from locally different sensitivity changes, as in Fig. 6.)

Adaptation adjusts to changes over time, while induction is a process that adjusts to changes across space [50]. In Fig. 7, the two gray squares have the same reflectance, yet the left square appears darker because it is surrounded by a lighter field. Thus the perceived lightness is again tied to the relative differences, or contrasts, within the image. These induction or simultaneous contrast effects are reminiscent of the contrast coding given by the center-surround receptive fields of retinal cells (cf. Fig. 1). A gray square that falls in the excitatory center of the field will lead to a larger response in such cells if the inhibitory surround is in darkness. Again, such adjustments could contribute to lightness constancy by factoring out the overall light level to preserve the response to the differences between the center and surround.

While such processes may begin at very peripheral stages of the visual system, lightness constancy also depends strongly on perceptual inferences that reflect much more sophisticated analyses of the stimulus structure. Figure 8 shows two striking examples of these inferences with regard to the perception of lightness and shading. In Fig. 8a, the squares labeled A and B have the same physical reflectance on the page, yet B appears much lighter [1]. This is because the cylinder is perceived to cast a shadow over square B. If less light is falling on B than A, yet both send the same intensity to the eye, then the B square must be made from material with a higher reflectance. And again, it is the surface reflectance, rather than the absolute light, to which our perception corresponds. In Fig. 8b, the circles that are darker on the bottom are perceived as round bumps that stick out of the page, while the circles darkened on the top appear as dents receding into the page. In this case, the visual system again interprets the luminance variations as shading on a three dimensional surface, and moreover, perceives the shading cues that would be consistent with lighting from above [39].

Checker-shadow illusion:
The squares marked A and B
are the same shade of gray.



Edward H. Adelson

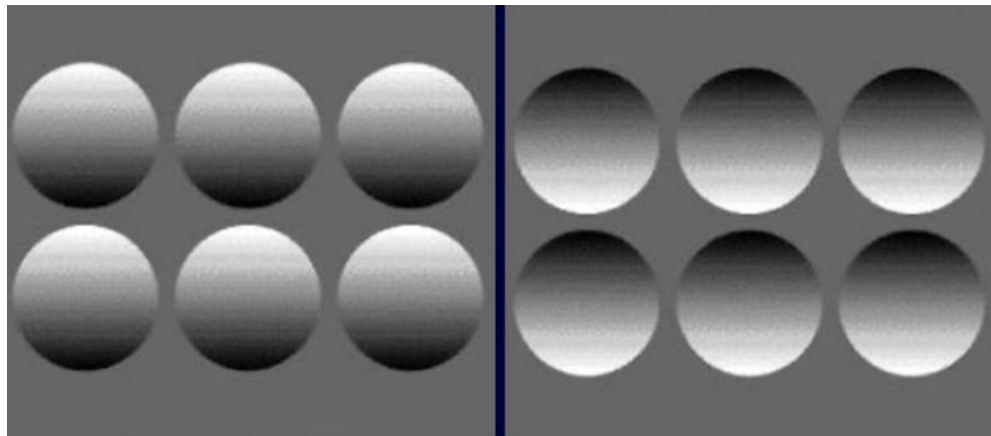


Fig. 8: Examples of high-level inferences in lightness perception. Top: the squares labeled A and B are physically the same, yet B appears lighter because it is perceived to be in shadow. Bottom: shading the circles on the bottom causes them to appear as solid bumps sticking out of the page, while adding the same shading to the top causes them to appear as dents receding into the page. Both perceptions arise from the assumption that the shading results from overhead lighting falling on a curved surface. Note that if the page is turned upside down then perception is reversed to again be consistent with lighting from above.

Color constancy

Achieving color constancy is more complex because the visual system has to solve for the spectral characteristics of the surfaces and the illuminants. Moreover, it has to do this by taking only three samples of each possible spectrum, i. e., the responses given by the L, M, and S cones. Color constancy necessarily fails under some conditions (e. g., when the illuminant is monochromatic) and is rarely if ever complete, yet in natural conditions the human visual system can exhibit a high degree of constancy [3]. This is in part because the spectral characteristics of natural illuminants and surfaces tend to vary gradually with wavelength, and because they tend to vary with few degrees of freedom. For example, the spectral variations among both daylight illuminants and many natural surfaces can be approximated by variations in a small number of basis functions [19, 30]. Thus three samples are often sufficient to closely approximate either a light or a surface.

However, because the color signal reaching the eye is the product of both the light and the surface, three samples do not suffice to represent both the illuminant and the object. To discount the illuminant therefore requires additional information or assumptions. For example, the illuminant could in principle be estimated from specular highlights in the image, or from the brightest surfaces (with the assumption that this surface reflects roughly evenly across the spectrum). A number of studies have explored the possibility that the visual system uses an assumption that the world is on average spectrally flat, or gray. The illuminant can then be estimated from the average chromaticity in the scene. To the extent that this assumption is valid, one way to approximate constancy is to adapt the cone sensitivities to adjust for the mean response. For example, we noted above that the gain of the cone signals is adjusted to compensate for a change in the overall light level. Suppose that the illumination shifts from flat to more energy at long wavelengths. The L cones will be strongly stimulated leading to a decrease in sensitivity, while the S cone will receive less light and thus increase in sensitivity. As a result, the average response across the cones will be similar under both illuminants, factoring out the mean color change and again largely preserving the relative differences between colors. Thus independent gain changes in each cone class, known as von Kries adaptation, provides a rough approximation to color constancy, and in fact is functionally equivalent to the process that Land invoked to account for his two-color projections. Viewing the slides through a red and white filter biased the average color toward red, yet if each cone class normalizes to the average then the mean perceived color remains achromatic, and stimuli that are less “red” than the average will take on the opposite hue. Von Kries adaptation alone cannot completely discount the illuminant, and would fail to maintain color appearance when the mean reflectance of the scene is not flat (as is often the case). Moreover, like lightness, perceived color depends on high-level inferences about the viewing geometry and material properties of scenes. These factors are currently an active area of research in color constancy [4].

Visual Experience

Thus far we have discussed how sensory information is encoded and how it is analyzed in order to extract ecologically important information about the world. However, none of this bears on the question of what it actually feels like to see, or the “sensory qualia” of visual awareness. The basis of conscious experience remains perhaps the central mystery of cognition and neuroscience, for despite active efforts there are still no obvious routes from neural activity to awareness – nor agreement on what awareness is – and as yet there are no objec-

tively verifiable measures of sensory qualia. That is, visual experience remains a private and subjective phenomenon, because we have access only to our own. As a result, the contents of our awareness are available only through introspection, yet much of the processes underlying perception precede awareness and are thus not accessible through introspection. Instead, to probe these processes vision science has had to develop an armament of methods that assess the ability to respond to stimuli rather than the experiences these stimuli elicit. A classic illustration of this point is that a pinhole or lens forms an inverted image, and this predicts that the image falling on our retina is upside down. Yet we experience the world right-side up. This problem perplexed such great minds as da Vinci, who developed elaborate models of the eye on the assumption that the image must be reinverted to accord with our experience. Yet the retinal image is in fact inverted. This conclusion is usually attributed to Kepler, and can be verified by again tapping the right side of your eye. This produces a visual sensation that appears to be coming from the left side of the world, because the pressure stimulates visual cells that are labeled to encode information about light from a particular direction. A second, modern example is the phenomenon of change blindness, in which observers fail to notice large physical changes introduced between two images of a scene [40]. Introspection suggests that we internalize a rich and detailed representation of the world, yet such studies reveal that at the level of awareness we instead encode only the gist of the scenes before us. A detail may become available only when we direct our attention to it.

Color Appearance

The phenomenology of color vision has played a central role in the philosophical and scientific debate over subjective experience, as illustrated by Locke's "inverted spectrum" [28]. The question he posed is how one could know whether two observers experience color in the same way. For example, the hues one experiences when looking at a rainbow might be inverted in the eyes of another, so that the sensation one feels as "blue" might feel "yellow" to the other. However, these experiences are private, and thus would appear overtly consistent as long as both observers use the same labels to describe a color. Thus a central question is what we can learn about color sensations from verbal reports.

The theory of color opponency was first advocated by Hering [14]. He observed that some colors are never perceived together, suggesting that they represent mutually exclusive sensations. For example, it is normally impossible to perceive a single light that appears both red and green or both blue and yellow. An opponent process that signaled either red or green depending on the sign of the response could readily account for such effects. This led to the still widely held notion that color appearance depends on two opponent mechanisms signaling red vs. green or blue vs. yellow sensations, and that the four "unique hues" that isolate these processes (e. g., a red that is not tinged with blue or yellow) are thus special because they represent the pure responses of these channels [17]. A bright-dark opponency forms a third axis. Color appearance can be ordered within a space defined by these three dimensions, with gray at the center (the neutral point for each opponent process), and can be specified by the three attributes of brightness (light-dark variations), hue (the relative amounts of each unique hue), and saturation (how pure or diluted the color appears). However, while the principle of opponency has been clearly validated, the actual processes identified by Hering have not. Specifically, the cone combinations that have been objectively measured in postreceptoral opponent channels are not the ones that give rise to pure red and green or blue and yellow

sensations, and at the level of visual cortex, color may be represented within many channels rather than the three classically supposed [46]. Thus the basis for the phenomenal primacy of the unique hues remains unknown.

Another approach to color experience has been to ask whether individuals who speak different languages label the spectrum in similar ways. Berlin and Kay found that languages contain only a small number of basic color terms used consistently by speakers (e. g., “red” but not “mauve”), and that these terms for different languages tend to point to similar regions of the spectrum [2]. This suggests that the way that humans categorize color probably does have a strong universal basis and that the categories are not arbitrarily shaped by language. However, it is not known whether basic color terms reflect the structure of the visual system or the structure of the visual environment (e. g., because some spectra are more common).

Perceptual Illusions

Despite the ambiguities in the retinal image and the many steps that the visual system must use to resolve these, perception itself seems effortless and immediate, and as we look about it is rare not to feel that the impressions formed are definite and correct (in the sense that they capture the ecologically important structure of the world). Visual illusions are striking precisely because they violate this sense of certainty about the validity of our perceptions. Illusions have always been a popular focus of perceptual studies, and provide important clues to the nature of perception. Yet surprisingly many are still poorly understood. It is also problematic even to try to define what constitutes an illusion. For example, the triangle in Fig. 4 appears as a phantom, but is arguably the best interpretation of the causal structure of the image.

Illusions can be roughly categorized according to the nature of the errors they represent [11]. Figure 9 illustrates three classic illusions which reflect different aspects of perceptual processing. The left drawing shows an ambiguous figure known as the Necker cube. The cube can be seen in two possible orientations, and which face appears closer can appear to randomly fluctuate over time. The ambiguity arises because the two interpretations are equally likely, and the fact that they alternate suggests that at some level the visual system entertains them both. Note, however, that it is only possible to perceive one perspective at a time. The processes of perception are probabilistic, but the product at any given moment is all or none. Figure 9b shows an example of an impossible figure, in which the prongs of the fork do not fit with the base. Such figures are important for demonstrating that conscious deduction plays little role in perceptual inference. Rather these inferences are part of the very structure of visual processing. Analyses that are applied locally are not constrained by the “force of logic” to be consistent globally. Finally, the drawing in Figure 9c shows the Muller–Lyer illusion, one of many examples of perceptual distortions. In this case the two central lines are the same length, yet when the flanks extend outward the line appears longer. Even this simple figure has evaded a definitive explanation, perhaps because it involves many factors [11]. For example, one account is that the length is judged at many spatial scales, and that mechanisms with low resolution will blur and thus extend the line with the outward flanks. However, another account is that we perceive the figures as perspective drawings. For example, the line with arrows can be seen as a corner projecting toward the observer, while the fans suggest a more distant corner. Since the two “walls” subtend the same angle then the closer one must be shorter.

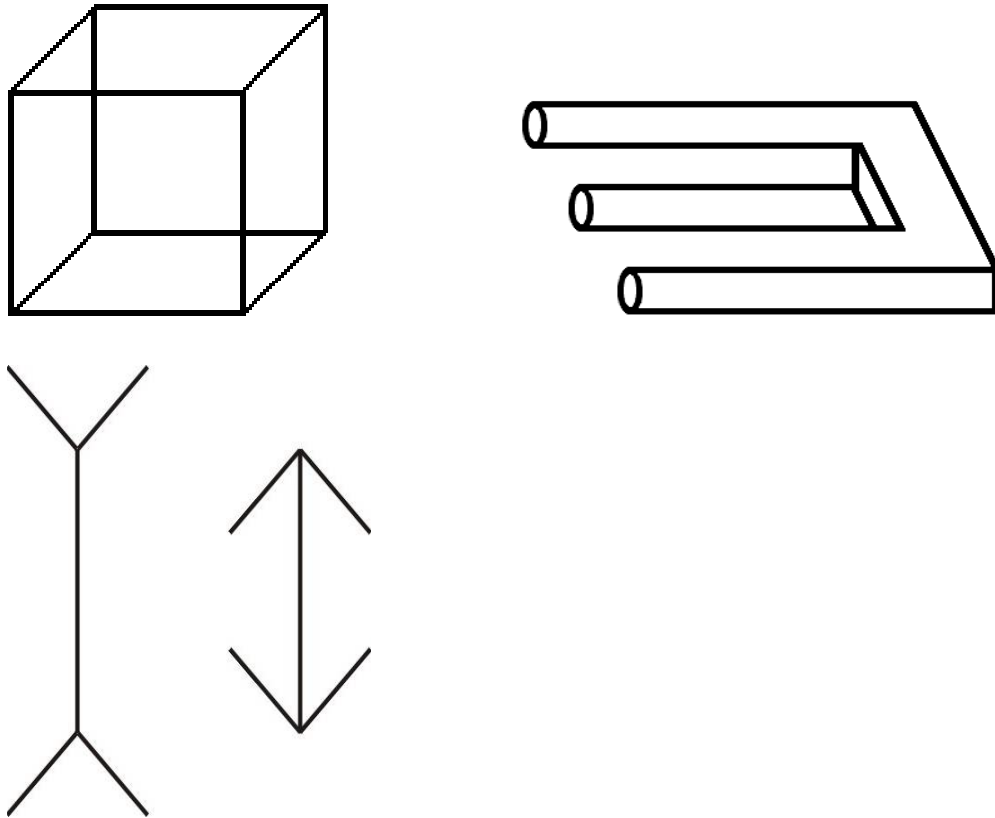


Fig. 9: Three classic examples of illusions. Top left: an ambiguous figure, the Necker cube, which oscillates between a cube facing up to the right or down to the left. Top right: an impossible figure of a fork, where the prongs and base are inconsistent because the figure on one end becomes the background on the other. Bottom: a perceptual distortion, the Muller–Lyer illusion, in which the two center lines are the same length yet the flanks cause the left line to look longer.

Most illusions involve situations contrived by clever artists, and it is much less common to feel fooled by nature. However, a spectacular exception to this is the moon illusion, in which the moon appears larger when it is near the horizon than directly overhead. Many theories abound to account for this, dating as far back as Ptolemy and continuing today [41]. One proposal is that the sky near the horizon appears farther away than the zenith sky because depth cues from nearby land features are available in one case but not the other. In order to subtend the same visual angle, the more distant horizon moon would therefore need to be larger. However, the paradox is that the rising moon not only looks larger but also closer, suggesting that more than one perceptual error might contribute to the impression. Like many aspects of perception, the illusion remains an intriguing puzzle.



References

- [1] E. H. Adelson, "Perceptual organization and the judgment of brightness", *Science* **262**, 2042 (1993).
- [2] B. Berlin and P. Kay, *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, 1969.
- [3] D. H. Brainard, "Color constancy", in *The Visual Neurosciences*, Vol. 2, pp. 948, L. M. Chalupa and J. S. Werner (eds.). MIT Press, Cambridge, 2003.
- [4] D. H. Brainard and L. T. Maloney, "Perception of color and material properties in complex scenes", *Journal of Vision* **4**, ii (2004).
- [5] G. S. Brindley, *Physiology of the Retina and the Visual Pathway*. Edward Arnold, London, 1960.
- [6] L. M. Chalupa and J. S. Werner, *The Visual Neurosciences*. MIT Press, Cambridge, 2003.
- [7] R. L. De Valois and K. K. De Valois, *Spatial Vision*. Oxford University Press, Oxford, 1988.
- [8] E. A. Deyoe and D. C. van Essen, "Concurrent streams in monkey visual cortex", *Trends in Neuroscience* **11**, 219 (1988).
- [9] A. K. Engel and W. Singer, "Temporal binding and the neural correlates of sensory awareness", *Trends in Cognitive Sciences* **5**, 16 (2001).
- [10] K. R. Gegenfurtner and L. T. Sharpe, *Color Vision: From Genes to Perception*. Cambridge University Press, Cambridge, 1999.
- [11] R. L. Gregory, *Eye and Brain*, 5th ed. Princeton University Press, Princeton, 1997.
- [12] Z. J. He and K. Nakayama, "Surfaces versus features in visual search", *Nature* **359**, 231 (1992).
- [13] H. v. Helmholtz, *Helmholtz's Treatise on Physiological Optics*. Dover, New York, 1866, 1911.
- [14] E. Hering, *Outlines of a theory of the light sense*. Harvard University Press, Cambridge, 1964.
- [15] I. Howard, *Seeing in Depth 1: Basic Mechanisms*. I Porteous, 2001.
- [16] D. H. Hubel, "Exploration of the primary visual cortex 1955–1978 (Nobel Lecture)", *Nature* **299**, 515 (1982).
- [17] L. M. Hurvich and D. Jameson, "An opponent-process theory of color vision", *Psychological Review* **64**, 384 (1957).
- [18] G. H. Jacobs, "The distribution and nature of colour vision among the mammals", *Biological Reviews* **68**, 413 (1993).
- [19] D. B. Judd, D. L. MacAdam, and G. Wyszecki, "Spectral distribution of typical daylight as a function of correlated color temperature", *J. Opt. Soc. Amer.* **54**, 1031 (1964).
- [20] P. Kaiser and R. M. B. Boynton, *Human Color Vision*. Optical Society of America, USA, 1996.
- [21] N. Kanwisher, "Domain specificity in face perception", *Nature Neuroscience* **3**, 759 (2000).
- [22] D. C. Knill and W. Richards, *Perception as Bayesian Inference*. Cambridge University Press, New York, 1996.
- [23] J. Krauskopf, D. R. Williams, and D. W. Heeley, "Cardinal directions of color space", *Vision Research* **22**, 1123 (1982).
- [24] S. W. Kuffler, "Discharge patterns and functional organization of mammalian retina", *Journal of Neurophysiology* **16**, 37 (1953).
- [25] E. H. Land, "Color vision and the natural image: Part 1", *Proc. Nat. Acad. Sci.* **45**, 115 (1959).
- [26] P. Lennie, J. Pokorny, and V. C. Smith, "Luminance", *J. Opt. Soc. Amer.* **A10**, 1283 (1993).
- [27] M. Livingstone and D. Hubel, "Segregation of form, color, movement, and depth: Anatomy, physiology, and perception", *Science* **240**, 740 (1988).
- [28] J. Locke, *An Essay Concerning Human Understanding*. Clarendon Press, Oxford, 1689, 1975.
- [29] D. I. A. MacLeod, "Colour discrimination, colour constancy and natural scene statistics", in *Normal and Defective Colour Vision*, J. D. Mollon, J. Pokorny, and K. Knoblauch (eds.). Oxford University Press, Oxford, 2003.
- [30] L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters", *J. Opt. Soc. Amer.* **A3**, 1673 (1986).

- [31] R. Mausfeld and D. Heyer, *Colour Perception: Mind and the Physical World*. Oxford University Press, Oxford, 2004.
- [32] J. D. Mollon, “‘Tho’ she kneel’d in that place where they grew...’”, *J. Exper. Biol.* **146**, 21 (1989).
- [33] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. Academic Press, San Diego, 2003.
- [34] J. Nathans, D. Thomas, and D. S. Hogness, “Molecular genetics of human color vision: The genes encoding blue, green, and red pigments”, *Science* **232**, 193 (1986).
- [35] I. Newton, *Opticks*, 4th ed. Dover, New York, 1730, 1952.
- [36] B. A. Olshausen and D. J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”, *Nature* **381**, 607 (1996).
- [37] E. Peterhans and R. von der Heydt, “Subjective contours: Bridging the gap between psychophysics and physiology”, *Trends in Neurosciences* **14**, 112 (1991).
- [38] S. Polyak, *The Vertebrate Visual System*. University of Chicago Press, Chicago, 1957.
- [39] V. S. Ramachandran, “Perceiving shape from shading”, *Sci. Am.* **259**, 58 (1988).
- [40] R. A. Rensink, “Change detection”, *Annu. Rev. Psychol.* **53**, 245 (2002).
- [41] H. Ross and C. Plug, *The Mystery of the Moon Illusion*. Oxford University Press, Oxford, 2002.
- [42] W. A. H. Rushton, “Pigments and signals in color vision”, *J. Physiol.* **220**, 1 (1972).
- [43] E. P. Simoncelli and B. A. Olshausen, “Natural image statistics and neural representation”, *Annu. Rev. Neurosci.* **24**, 1193 (2001).
- [44] A. Stockman and L. T. Sharpe, “Cone spectral sensitivities and color matching”, in *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe (eds.). Cambridge University Press, Cambridge, 1999.
- [45] L. G. Ungerleider and M. Mishkin, “Two cortical visual systems”, in *The Analysis of Visual Behavior*, D. J. Ingle, R. J. W. Mansfield and M. S. Goodale (eds.). MIT Press, Cambridge, 1982.
- [46] M. A. Webster, “Human colour perception and its adaptation”, *Network: Computation in Neural Systems* **7**, 587 (1996).
- [47] M. A. Webster, “Pattern selective adaptation in color and form perception”, in *The Visual Neurosciences*, Vol. 2, p. 936, L. M. Chalupa and J. S. Werner (eds.). MIT Press, Cambridge, 2003.
- [48] H. R. Wilson and F. Wilkinson, “Spatial channels in vision and spatial pooling”, in *The Visual Neurosciences*, Vol. 2, L. M. Chalupa and J. S. Werner (eds.). MIT Press, Cambridge, 2003.
- [49] G. Wyszecki and W. S. Stiles, *Color Science*, 2nd ed. Wiley, New York, 1982.
- [50] Q. Zaidi, “Color and brightness induction: from Mach bands to three-dimensional configurations”, in *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe (eds.). Cambridge University Press, Cambridge, 1999.