Lightness and Retinex Theory

Article *in* Journal of the Optical Society of America · February 1971

DOI: 10.1364/JOSA.61.000001 · Source: PubMed

John J McCann
McCann Imaging

Reprinted from JOURNAL OF THE OPTICAL SOCIETY OF AMERICA, Vol. 61, No. 1, 1–11, January 1971
Printed in U. S. A.

Lightness and Retinex Theory

EDWIN H. LAND* AND JOHN J. MCCANN
Polaroid Corporation, Cambridge, Massachusetts 02139
(Received 8 September 1970)

Sensations of color show a strong correlation with reflectance, even though the amount of visible light reaching the eye depends on the product of reflectance and illumination. The visual system must achieve this remarkable result by a scheme that does not measure flux. Such a scheme is described as the basis of retinex theory. This theory assumes that there are three independent cone systems, each starting with a set of receptors peaking, respectively, in the long-, middle-, and short-wavelength regions of the visible spectrum. Each system forms a separate image of the world in terms of lightness that shows a strong correlation with reflectance within its particular band of wavelengths. These images are not mixed, but rather are compared to generate color sensations. The problem then becomes how the lightness of areas in these separate images can be independent of flux. This article describes the mathematics of a lightness scheme that generates lightness numbers, the biologic correlate of reflectance, independent of the flux from objects.
INDEX HEADINGS: Vision; Color.

Most of us assume that, subject to a variety of compensatory factors, we see in terms of the amount of the light coming from objects to our eye; we think that in a particular scene there is more light coming from white objects than from black objects; we think there is more long-wave light (so-called red light) coming from red objects than from blue objects.

Yet, when we measure the amounts of light in the world around us, or when we create artificial worlds in the laboratory, we find that there is no predictable relationship between flux at various wavelengths and the color sensations associated with objects. Accordingly, we believe that the eye must have evolved a system which, though using light as the communication medium with the world, has become as nearly independent of energy as is biophysically possible. In short, color sensations must be dependent on some as yet undefined characteristic of the field of view, a characteristic that can be communicated to us by the light with which we see, even though the amount and composition of the light are everywhere variable and unpredictable; the eye must have evolved around such a permanent characteristic of the field of view. This paper describes our search for that characteristic.

A major visual phenomenon is that objects with low reflectance look dark, objects with high reflectance look light, objects with reflectance higher in the long-wave portion of the spectrum than in the short-wave look reddish, objects with reflectance higher in the short-wave portion than in the long-wave look bluish, and so on. It is with reflectance that sensation of color is strongly correlated when we view the world around us.[1–8] Yet ascertaining reflectance in any of the familiar ways requires an operational step which the eye cannot take. For example, the eye cannot insert a comparison standard next to the object which it is regarding.

Furthermore, what reaches the eye from each point is clearly the product of the reflectance and the illumination. The illumination from the sun is modulated by clouds, atmosphere, water, mountains, trees, houses, etc. As every photographer knows, the sun and sky produce every conceivable combination of sunlight and skylight. Even less uniform illumination is provided by artificial light because the distance from the light source drastically affects the illumination falling on any point.

We are then left with the circular logical problem that, because the light coming to our eye is the product of the reflectance and illuminance, our eye could not determine reflectance unless the illuminance is uniform and the eye could not determine illuminance unless the reflectance is uniform. In general, across the field

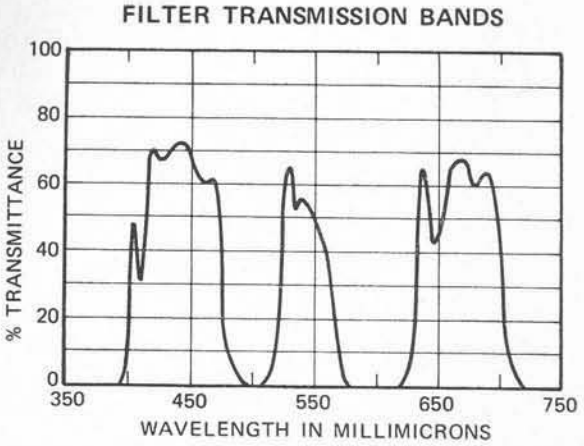


FIG. 1. Spectral transmittances of bandpass interference filters.

of view, neither reflectance nor illuminance is known; and neither is uniform.

INDEPENDENCE OF COLOR SENSATION FROM FLUX–WAVELENGTH DISTRIBUTIONS

To demonstrate the extent to which color sensation is independent of flux–wavelength distribution, we will describe a simple quantitative laboratory experiment. An extended array of rectangular, colored papers is arranged to look like the paintings of the artist Mondrian.[9,10] To reduce the role of specular reflectance, the papers are not only matte, but are also selected to have a minimum reflectance as high as or higher than 10% for any part of the visual spectrum. The Mondrian-like pattern is illuminated by three illuminating projectors with sharp-cut bandpass filters,[11] one passing long waves, one middle-length waves, and one short waves (Fig. 1). The flux from each of the three projectors is changed by a separate variable transformer. The filters are selected on two bases: first, to minimize the diversity of color sensations from the array of colored papers when only one projector is turned on; second, while satisfying the first condition, to transmit as wide a band of wavelengths, and as much light as possible.

With all three illuminating projectors turned on, the variable transformers are set so that the whole array of variegated papers is deeply colored, and so that, at the same time, the whites are good whites. This is not a critical setting. Then, using one projector at a time, and hence only one waveband at a time, we measure with a telescopic photometer the luminance at the eye from any particular area, say a white rectangle. Thus, we obtain from a white rectangle three numbers that are proportional to the three luminances at the particular location of the eye. (The subsequent procedures constitute a null experiment. The radiance-vs-luminance function, the particular units of measure, the wavelength sensitivity, and the linearity of the meter are not significant in the experiment.)

In one example, the readings from a white area were 6 long-wave units, 35 middle-wave units, and 60 short-wave units. We turned the photometer to another area, such as a dark brown. We then separately adjusted the transformers to settings such that the three luminances at the eye were 6, 35, 60. Thus the luminances from the new area were identical to the three luminances previously reaching the eye from the first rectangle. The color sensation from the new area remained essentially unchanged (dark brown) despite the fact that the wavelength–luminance composition for that area had changed from whatever it might have been to 6, 35, 60. We then pointed the photometer towards a series of different areas: bright yellow, blue, gray, lime green, and red. The illumination of each area was readjusted, in turn, so that the three luminances coming to the observer's eye were 6, 35, 60. After each of the new illuminations was adjusted so that the photometer read 6, 35, 60 for the long, middle, and short wavelengths, each area appeared essentially unchanged. Thus, the observers reported that the color sensations from the series were yellow, blue, gray, green, and red. When the variable transformers were changed in this way to produce the standard set of three luminances for any square, then all the other areas nevertheless continued to generate their original color sensation (although in a few areas there were some slight changes). Dramatically, the retention of the color sensations was related to the reflectances of the papers—not to the product of reflectance times illumination, although this product appears to be the only information reaching the eye.

Therefore, the color sensations in the display have a completely arbitrary relation to the composition of light in terms of wavelength and luminance of any one point. The luminance-vs-wavelength distribution of each object in the world around us cannot tell us whether an object is white, gray, or black; the ratio of fluxes at various wavelengths cannot determine whether a point on an object is reddish, greenish, bluish, or grayish. The mystery then is how we can all agree with such precision about blacks, whites, grays, reds, greens, browns, yellows, when there is no obvious physical quantity with which to describe how we know at all the color of the objects we are seeing.

It might occur to the reader that such a large change of relative luminance, a change such as we produce by altering the output of the long-wave projector relative to other projectors, is countered by a compensatory adaptation in the eye. If, in the previous experiments, changes of adaptation compensated for the changes of flux coming to the eye, then deliberately causing changes of adaptation should have a significant effect on the color appearance of objects. To produce an extremely large difference between the state of adaptation to long waves and the state of adaptation to middle and short waves, we asked observers to wear deep-red, dark-adaptation goggles, described by Hecht,[12] for $\frac{1}{2}$ h, to allow maximum regeneration of middle- and short-wave visual pigments. In order to insure an ample domain for adaptation, the level of illumination of the display was maintained at a sufficiently high level. (The white areas had luminances for the middle- and short-wave bands between 100 and 1000 times higher than the threshold for cone response after 30 min in the dark.) When the observers removed the goggles, they reported at the first instant, as well as later, that the colors of the paper squares in the Mondrian were essentially unchanged. The experiment was repeated with the deep-red, dark-adaptation filter over only one eye. At the end of the adaptation period, the observers, using the binocular-comparison technique, reported slight shifts of the color sensations but none so large as to change the color names. (Indeed, in our theory a change of photochemical adaptation is unimportant, for the same reasons that a change of the flux of one of the illuminants is unimportant. Similarly, reasonable variations of the native concentrations of visual pigments are not important from time to time, or from individual to individual.)

In another set of adaptation-related experiments the 6, 35, 60 Mondrian experiments were repeated with the observers seeing the Mondrian for less than 1/100 s. The experimental procedure was exactly the same as in the first 6, 35, 60 Mondrian experiments, with the exception that the observers looked at the Mondrian through a photographic shutter. The projectors were set so that the long-, middle-, and short-wave luminances from a white area were 6, 35, 60 and the observers reported that the area appeared white. The projectors were then set so that other areas had luminances of 6, 35, 60 and the observers reported that these areas, as before, produced sensations of brown, yellow, blue, gray, green, and red.

These experiments are significant because they show that the visual system uses a processing mechanism that is not merely independent, but instantaneously independent of the wavelength–luminance composition of the light coming to the eye. These mechanisms are not controlled by processes that are time dependent, such as the changes of the visual pigments that are due to differences of duration or intensities of adapting illumination.

If a particular rectangle is moved to various positions in the Mondrian, where it is surrounded by new sets of colored rectangles, the color sensation does not change significantly. The color sensation depends only on the long-, middle-, and short-wave reflectances of the rectangle and not on the properties of the neighboring rectangles. This independence of the neighboring rectangles holds for all flux settings of the illuminating projectors.

Because all these experiments, which show that any given wavelength–luminance combination, within limits as wide as the reflectance variations of these papers,

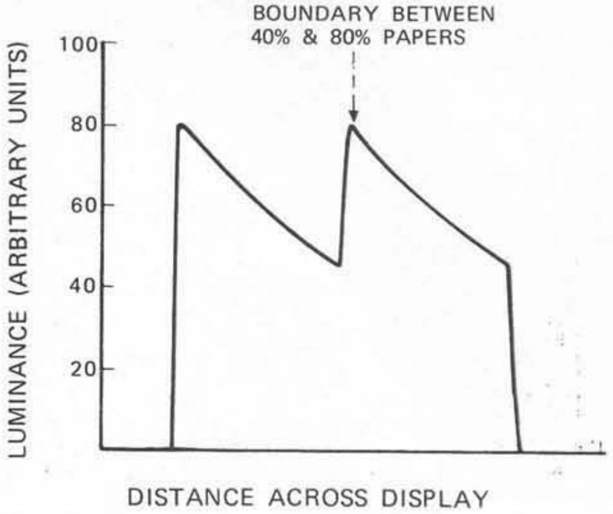


FIG. 2. Luminance vs position for two-squares-and-a-happening experiment.

can produce any color sensation, and because of many other kinds of laboratory experiments,[13–17] we came to the conclusion that a color sensation involves the interaction of at least three (or four) retinal–cortical systems. Each retinal system starts with a set of receptors peaking, respectively, within the long-, middle-, or short-wave portion of the visible spectrum. Each system forms a separate image of the world; the images are not mixed but are compared. Each system must discover independently, in spite of the variation and unknowability of the illumination, the reflectances for the band of wavelengths to which that system responds.

We invented a name, retinex, for each of these systems. A retinex employs as much of the structure and function of the retina and cortex as is necessary for producing an image in terms of a correlate of reflectance for a band of wavelengths, an image as nearly independent of flux as is biologically possible.

It is convenient to refer to the differences in this image as steps of lightness,[18–20] the whites being called light, and the blacks being called dark. Unfortunately, as Evans[21] points out, dark is also used to describe the quite different family of experiences associated with change of illumination. Nevertheless, following him, we shall call the steps in the scale from black to white, steps of lightness. In our theory, it is an image in terms of lightness, which is produced by each retinex for the portion of the spectrum to which its pigment responds.

The color sensation for any area is determined by the three lightnesses that are arrived at independently by the three retinexes. Because the lightnesses of an area are here defined as the biologic correlates of three reflectances, it follows within this conceptual framework that the color sensation is not dependent on illumination or flux, but on reflectance. Our original problem is converted into a new one: How does each retinex generate for each area the appropriate light-

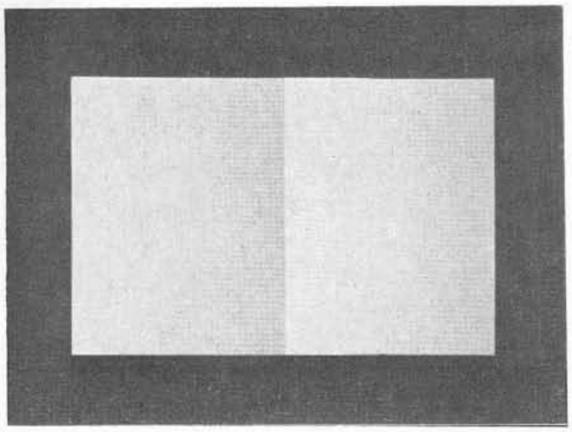


Fig. 3. Picture of two-squares-and-a-happening experiment. Place a pencil over the boundary between the two gray areas.

ness, the biologic correlate of reflectance that is independent of illuminance?

The scheme that we are about to describe for answering this question is one of a number of approaches that we have been investigating. All these schemes are designed to solve the same problem, namely: For one retinex, given the flux from each point in an entire scene, and assuming that nothing is known about the pattern of illumination and nothing is known about the reflectances, how can the biologic system generate a set of values that we experience as lightness? The particular scheme we will describe is the first that we have found to satisfy our criteria.

EDGES

The experiment that we call two squares and a happening provides striking evidence of edges as the source of lightness information. A piece of paper that reflects 80% of the light that falls on it is placed to the right of a piece of paper that reflects 40% of the light. A fluorescent tube is mounted in front and to the left of the papers. The tube is carefully positioned so that twice as much light falls on the center of the 40% paper as falls on the center of the 80% paper. The light, being a line source, produces an approximately linear gradient across the papers, and the reflected luminances at each corresponding point of the two papers are equal. The graph of luminance vs position on the display is shown in Fig. 2. The 40% paper on the left looks darker than the 80% paper on the right. Figure 3 is a photograph of the experiment.

What increases our interest is that when a long narrow object, a happening, obstructs the boundary between the left and right areas, the two areas are then perceived as having the same lightness. Long narrow strips of colored papers in parallel, or three-dimensional objects such as a pencil or a piece of yarn, make the two areas change from looking uniform and different to looking uniform and indistinguishable; yet, the only alteration of the display is the obscuration of the edge. We can see this by placing a pencil on the boundary between the areas in Fig. 3.

The experiment was important in the development of our ideas of how the visual system could generate lightnesses. The fact that obscuring the edge information could change the appearance of these areas meant that the edges are a very important source of information. It suggested that the change of luminance at the junction between areas both constituted an edge and also led to the visual difference between the whole two areas. The word edge suggests a sharp, in-focus boundary. Experiments, however, show that the sharpness or focus of the boundary is not at all critical. For example, Fig. 3 can be viewed through optometric lenses to change the boundaries from being sharp and in focus to a variety of fuzzy out-of-focus stages. Areas with boundaries quite out of focus look essentially the same as when they are in sharp focus.

What mechanism can we imagine that would discover edges and characterize adjacent areas in a way consistent with our experiences with the happening, a mechanism that will also discover the reflectances in the Mondrian even when it is in nonuniform illumination?

Let us imagine two light detectors placed to measure the luminance from two different places on a piece of paper. If the illumination is nonuniform, then the luminances at these two positions will, of course, be different. When the two detectors are placed closer and closer together, the luminances approach the same value and the ratio of the outputs approaches unity. This will be true of almost any two adjacent points. However, if the two detectors bridge the boundary between two areas of differing reflectance, then the ratio of the outputs of these detectors will approach the ratio of the reflectances. Thus, the simple procedure of taking the ratio between two adjacent points can both detect an edge and eliminate the effect of nonuniform illumination. Processing the entire image in terms of the ratios of luminances at closely adjacent points generates dimensionless numbers that are independent of illumination. As the distance between detectors is decreased, each number approaches a limit equal to the ratio of the reflectances, the reflectances themselves having not yet been ascertained.

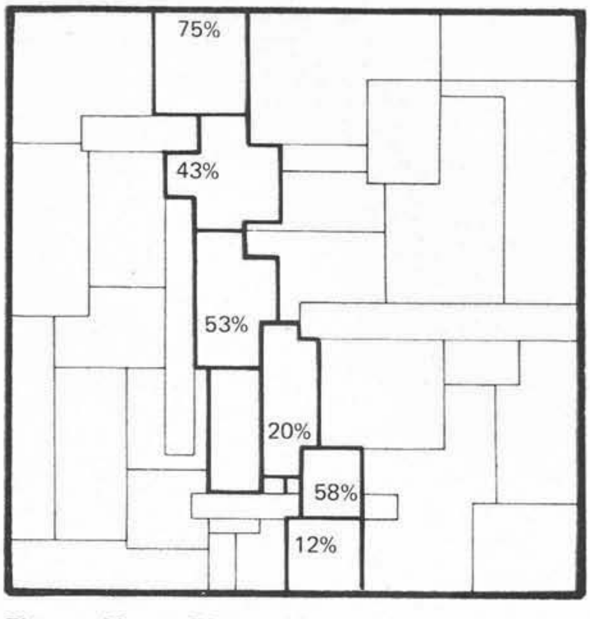
ENTIRE FIELD OF VIEW

Given a procedure for determining the ratio of reflectances between adjacent areas, the next problem is to obtain the ratio of reflectances between any two widely separated areas in an entire scene. We solve the problem in the following way: Find the ratio of luminances at the edge between a first and a second area, and multiply this by the ratio of luminances at the edge between the second and a third area. This product of sequential ratios approaches the ratio of reflectances between the first and third areas, regardless of the distribution of illumination. Similarly, we can obtain the ratio of reflectances of any two areas in an image, however remote they are from each other, by multiplying the ratios at all the boundaries between the starting area and the remote area. We can also establish the ratio of the reflectance of any area on the path to the reflectance of the first area by tapping off the sequential product at that area.

Consider a Mondrian similar to the colored one in complexity and randomness, but consisting of black, gray, and white papers (see bound transparency, Fig. 4). That is, in this Mondrian each piece of paper has the same reflectance for all wavelengths. The reflectances of each area along one path between the top and the bottom are shown in Fig. 5. If we apply the sequential-multiplication technique to these reflectances, we can determine the ratio of the top reflectance to the bottom reflectance, as shown in Fig. 5. Note that the number we get by sequential multiplication, 75/12, equals the number we would get if the bottom area were contiguous to the top area and we took the ratio of their luminances. We are now coming close to the solution of the problem that we defined at the beginning of our discussion. How can the eye ascertain the reflectance of an area without, in effect, placing a comparison standard next to the area? The sequential product can be used as a substitute for the placement of two areas adjacent to each other, thus defining a photometric operation conceptually feasible for the eye.

We placed a fluorescent tube to illuminate the Mondrian from below so that more light fell on the bottom of the display than on the top. We adjusted the position of the light so that exactly the same luminance was coming to the eye from a high-reflectance area at the top of the display and a low-reflectance area near the bottom. If the luminance determined the lightness of an area, the low-reflectance area and the high-reflectance area should look essentially alike; in fact, they do not. Although the luminances of the two areas are equal, the high-reflectance area at the top looks dramatically lighter than the low-reflectance area at the bottom (see areas indicated by arrows in bound transparency, Fig. 6).[22] Clearly, the visual processes that determine the lightness of an area are not governed by the luminance of that area.

Figure 7 shows the luminances along a path from the top of the Mondrian to the bottom. Note that the luminance at the center of the top area is the same as the luminance at the center of the bottom area. Considering the top area alone, note that the luminance (in arbitrary units) increases from 118 at its center to 140 at its lower edge. The ratio between the bottom edge of the first area and the top edge of the second area is 140 to 80. The luminance of the second area increases from 80 to 115 from upper edge to lower edge. The ratio of the second area to the third is 115 to 150. As we continue down the path, we obtain the ratios shown at the bottom of Fig. 7. The product of all the



$$\frac{75}{43} \times \frac{43}{53} \times \frac{53}{20} \times \frac{20}{58} \times \frac{58}{12} = \frac{75}{12} = \frac{6.25}{1}$$

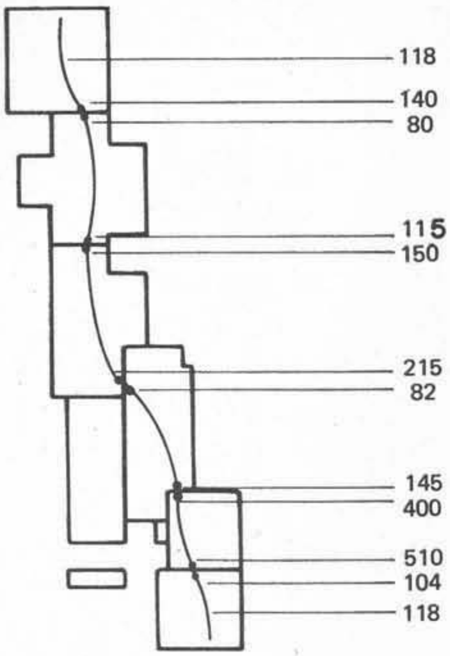
FIG. 5. Reflectance along one path between the top and bottom of a black-and-white Mondrian. The numbers at the bottom indicate the ratios of reflectances at adjacent edges along the path.

ratios along the path from the high-reflectance area at the top to the low-reflectance area at the bottom is 6.25. This number is equal to the ratio of reflectances of the top and bottom areas. Thus, without determining the reflectances and without determining the illumination, however it varies, we have determined a number exactly equal to the ratio of reflectances of these two areas. Yet the two areas have the same luminance as each other and are remote from each other by the whole width of the display. Furthermore, this procedure of sequential multiplication of edge ratios can generate values equivalent to relative reflectance for all areas along the path.

CONSISTENCY OF SEQUENTIAL PRODUCTS ON DIFFERENT PATHS

Let a number of different paths start from a given area and wander back and forth over the display, all to arrive finally at a distant area, which we wish to evaluate with respect to the starting area. If we compute the sequential products along each of these paths, we obtain the ratio of reflectances of the remote area to the starting area for each path. In this case the starting and remote areas for all the paths are the same, therefore the terminal sequential products are identical.

If, instead of having all the paths start from a single area, the paths start from different areas, wander over the display, and all terminate in a single remote



$$\frac{140}{80} \times \frac{115}{150} \times \frac{215}{82} \times \frac{145}{400} \times \frac{510}{104} = \frac{6.25}{1}$$

FIG. 7. Luminances of Mondrian (illuminated from below) at particular points along the path from top to bottom. The numbers at the bottom indicate the ratios of luminances at adjacent edges along the path.

area, then once again the sequential products give the ratios of reflectances of the starting areas to the single terminating area. However, because all the starting areas had different reflectances, the sequential products at the single distant area will be different from each other. Our ultimate purpose is to describe any area by relating its reflectance to a single, standard, high reflectance somewhere in the Mondrian or to several equally high reflectances. Therefore, we require at this stage of our analysis that all the paths start from areas having the same reflectance. If the value of the starting reflectance is 100% (by some standard), then the sequential product for any area in the display will be numerically equal to the reflectance of the area (related to the same standard). Therefore, we approach the problem of locating a standard area by seeking the area of highest reflectance. A variety of operations can be used.

One that seems simple, but is not, is to scan the entire scene to find the area or areas with the highest reflectance. This technique requires two separate operations: finding the highest reflectance and then computing the sequential products. Now finding the highest reflectance would require taking all the sequential products, comparing the results, selecting the highest value, and then starting all over again to determine the reflectance relative to the area that had this highest value. This second step requires that the information about the areas of highest reflectance be sent back to the receptors, from wherever the operation of comparing and selecting takes place, and a subsequent procedure such that every path starts with an area thus identified as having the highest reflectance. Although this technique is mathematically valid we feel that it is not readily transposed into biological mechanisms.

We therefore sought a technique that can automatically establish the highest reflectance without a separate first scanning step. We adopted the convention that the initial ratio is the ratio of the signal of the second receptor to that of the first (followed by the third to the second, etc.). Then, regardless of the true reflectance of an area, our technique supposes that the first receptor in any path is reporting from an area that reflects 100% of the light. Because of the deliberately adopted fiction that the starting area reflects 100%, irrespective of its real reflectance, the sequential product becomes greater than 1.0 whenever the path reaches an area whose reflectance is higher than that of the starting area. Attainment of a sequential product greater than 1.0 indicates that the sequence should be started afresh and that this new, higher reflectance should be next supposed to be 100%. Note the very important condition that the sequence is not started afresh by a single ratio that is greater than 1.0, but only when the sequential product for the whole chain to that point becomes greater than 1.0. This distinction is at the heart of the technique for finding the highest reflectance in the path. As the path proceeds, the sequential product always starts over at unity when the path encounters an area with a reflectance higher than the highest previously encountered. We will discuss later the role of sequential products tapped off before the final reset to unity at the highest reflectance on a path.

A PHYSICAL MODEL

Imagine a long fibrous path, like a wire of some kind, on which pairs of photoreceptors are mounted close together. These pairs do not know where the edges fall. They take the ratio at adjacent points along the wire. Most of the ratios will approach 1/1, indicating no change. In the immediately following discussion, let us assume that any ratio that approaches 1.0 is exactly equal to 1.0. Later we will discuss how this can be acceptable. When a pair of photoreceptors happens to straddle an edge, they indicate a larger or smaller ratio. If the pattern is lit from its bottom and even if there is 10 times as much light falling on the bottom as on the top, the close spacing of the photoreceptors insures that, at any given point, the difference of illumination will be small. Every bridge pair of photocells will read the ratio of adjacent points as 1.0 until an edge intervenes. The bridge pairs of photocells, not knowing anything about edges, sometimes generate ratios equal to 1.0 and other times generate larger or smaller ratios.

Figure 8 is a specific example of how the machine operates. The pattern consists of a series of papers that reflect 60, 20, 40, 100, 60, 80, and 30% of the light falling on them. Let us assume, for the moment, that the illumination is uniform, so that we can assign one value per area rather than many different luminances. If we set the uniform illuminance equal to 100 in arbitrary units, the luminance will be equal to the percent reflectance of the papers. This facilitates comparison of the output of the machine with the reflectance of the display.

The members of the first pair of detectors read 60 and 20; as a bridge, they read the ratio as 20/60 or 0.33. This means that the machine, having assigned to the first area a lightness equivalent to 100% reflectance, reads at the second area 33%. These values do not correspond with the actual reflectances but are proportionately correct. The next pair lies within the boundaries of an area and has a ratio of 20/20 or 1.0. Thus, multiplying by 1.0 transmits the edge-ratio signal across an area. We shall see later the benefits of treating the system logarithmically; then, for pairs of receptors between boundaries the log of 1.0, namely 0, is added to the log sequential product. The next reading on the path is 40/20, which when multiplied by 20/60 equals 40/60 or 0.67, corresponding to a proportionately correct, but absolutely incorrect, reflectance of 67%.

The operation of the next pair of receptors shows how the system automatically finds the highest reflectance and restarts there. The edge ratio from this pair is 100/40. We have already had fractions from edge ratios that are larger and smaller than 1.0. However, in this case, 100/40 times the continued product 40/60 equals 100/60 and this is the first sequential product that has a value greater than 1.0. The policy that we have established is that any value greater than 1.0 will be reset to equal 1.0. For all fresh starts, the assigned value is 100%; thus the output with the sequential product 100/60 will also be 100%. Once again, throughout the rest of the string of receptors, the sequential product will establish the reflectances relative to the reflectance of the area at which the sequential product was reset to 100%. But this 100% is either a real 100% reflectance or, being the highest value on this whole path, is a permanent substitute for the real 100% on this path. There will be no further fresh starts, and the reflectances are from that point on as close as they are going to be to real reflectance.

What do we do with the areas between the start and the first maximum reflectance? If we think about the properties of a long path that closes on itself, it becomes clear that the highest reflectance on the chain will finally dominate the chain, if we assume that the image is stationary long enough for the system to reach a steady state.

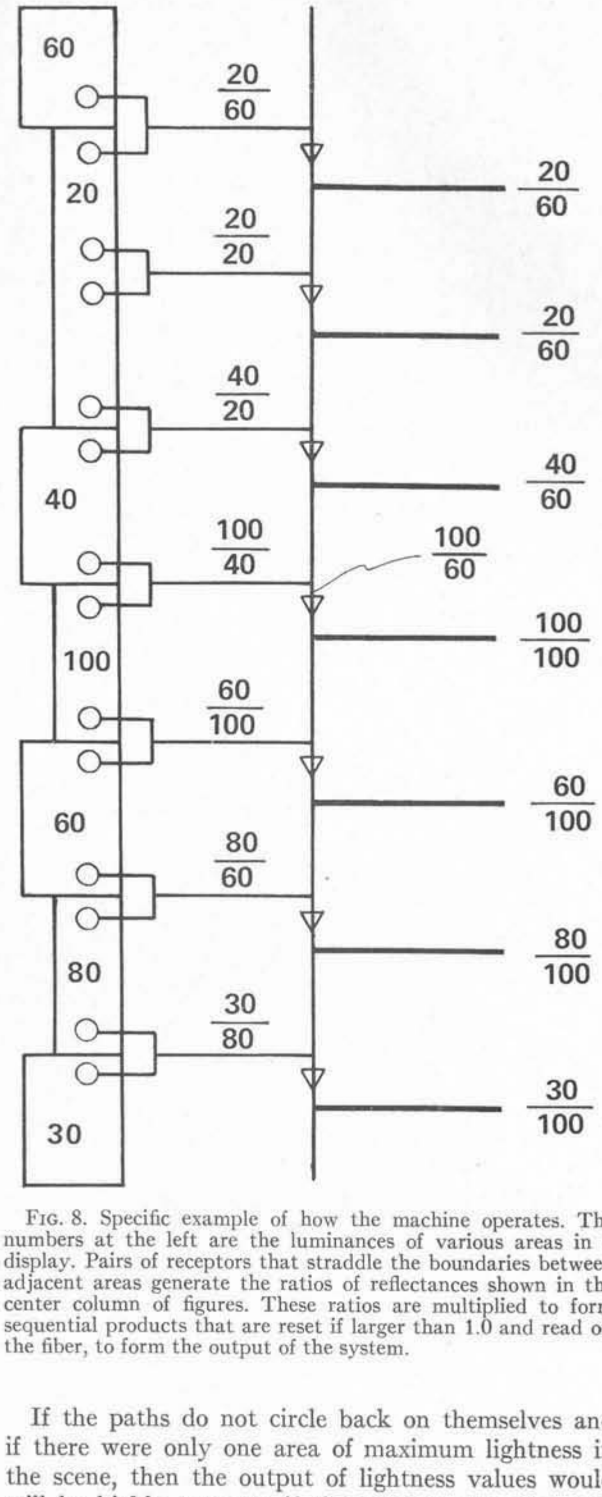


FIG. 8. Specific example of how the machine operates. The numbers at the left are the luminances of various areas in a display. Pairs of receptors that straddle the boundaries between adjacent areas generate the ratios of reflectances shown in the center column of figures. These ratios are multiplied to form sequential products that are reset if larger than 1.0 and read off the fiber, to form the output of the system.

If the paths do not circle back on themselves and if there were only one area of maximum lightness in the scene, then the output of lightness values would still be highly accurate if the number of paths were sufficiently large. Let us examine the properties of two paths going in opposite directions. If we consider the example in Fig. 8, the first path would have percentage outputs of 100, 33, 67, 100, 60, 80, and 30. A path in the opposite direction would have 60, 20, 40, 100, 75, 100, and 100. Thus averaging only two paths gives

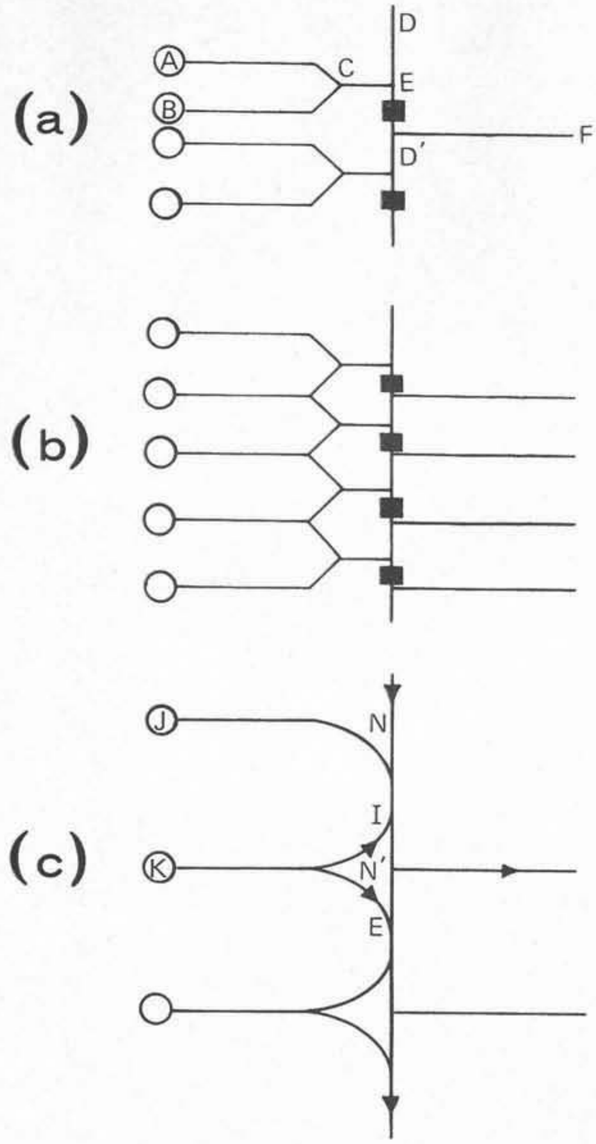


FIG. 9. A variety of equivalent sequential-products models. In (a), two opposed logarithmic receptors (A,B) first sum with each other (C) and then sum with the continued product (D at E). This total quantity is both the readout of the system (F) and the new continued product (D') that is combined with the next receptor pair output. In (b), each photocell is the leading photocell for one bridge pair and the trailing photocell for the next bridge pair. In (c), a third variation is perhaps more biologically oriented. The receptor K transmits its signal to its synapses I and E. Synapse I is an inhibitory synapse and adds to the sum of J and the sequential product N. The new sequential product is formed at L and is tapped off the chain between the two synapses I and E. Synapse E is excitatory and combines with this new sequential product N' for the computation of the next sequential product.

the values 80, 27, 53, 100, 68, 90, and 65 compared to the actual reflectances 60, 20, 40, 100, 60, 80, and 30. The average of the two paths is more accurate than either of the separate paths.

For convenience, these schematic examples have used uniform illumination and luminance numerically equal to reflectances. Because the members of each bridge are close to each other, the readings of the bridge pairs on a pathway will not change significantly as the illumination is made nonuniform. The machine will continue to read approximate reflectances and will be independent of luminance: It is this competence that is the essential purpose of the machine.

If we were to build an electronic model, we would probably make some changes for practical reasons. The most obvious change would be to use logarithmic receptors so that any multiplication could be performed by the summation of positive and negative voltages or excitatory and inhibitory responses. Figure 9(a) shows a simplified conception of the scheme. Two opposed logarithmic receptors (A, B) first sum with each other (C) and then sum with the continued product (D at E). When the logarithm of the sequential product is greater than 0, then the path restarts and the logarithm of the sequential product is reset to 0. This operation is indicated in the diagram by the black square below E. This total quantity is both the readout of the system (F) and the logarithm of the continued product (D') that is combined with the output of the next receptor pair. Furthermore, a logarithmic system has properties that assist in implementing the scheme of resetting the chain where the advent of a new high yields a sequential ratio greater than unity. Because the logarithm

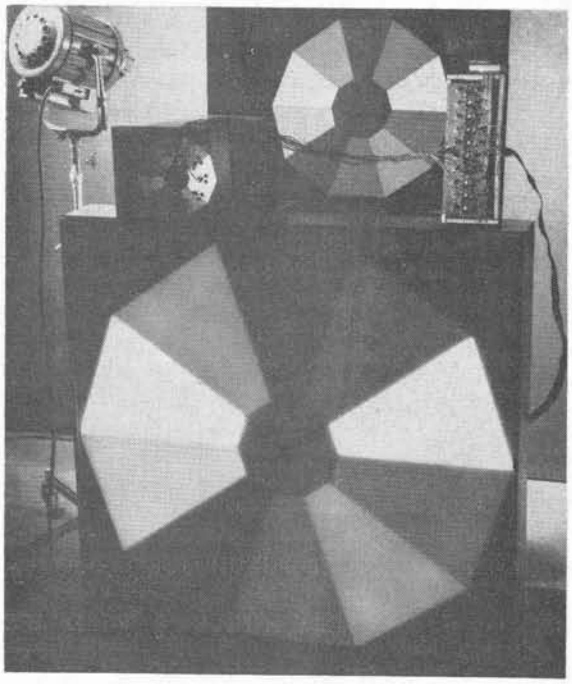


FIG. 10. Photograph of retinex machine reproducing the white, gray, and black wheel. The spotlight on the far left illuminates the wheel on the back wall and the camera on the center left forms an inverted image of it. The photocell pairs in the camera send the ratios of luminances to the electronics on the right which computes the sequential product and transmits it to the display below. The machine gives the same outputs regardless of the position of the spotlight.

of 1.0 is 0, we use an electrical system that is negative only and that will reset any positive potential to 0 voltage, thus limiting the sequential product.

Figure 10 shows a photograph of the retinex machine. It consists of a camera, with photocells in the film plane, electronics for computing the sequential product, and a unit to display the sequential products. The white, gray, black wheel mounted on the back wall is illuminated by the spotlight on the left. The inverted image of the wheel can be seen in the ground glass of the camera on the left. Bridge pairs of photocells are arranged on the ground glass so that each pair straddles an edge in this image. The pairs of photocells are silhouetted on the ground glass. The output of the photocells is processed by the electronics to produce a sequential product for each area in the wheel. The wires going from the electronics to the display in the foreground transmit the sequential product to the lamps mounted in boxes behind the ground-glass panel. The lightnesses of these boxes, when the machine is operating, are the correlates of the original reflectances. No matter how drastically the illumination is changed on the original papers in the wheel, the lightnesses in the final display will be independent of the changes of illumination and hence independent of luminance. Figure 11 shows a schematic diagram of the machine.[23]

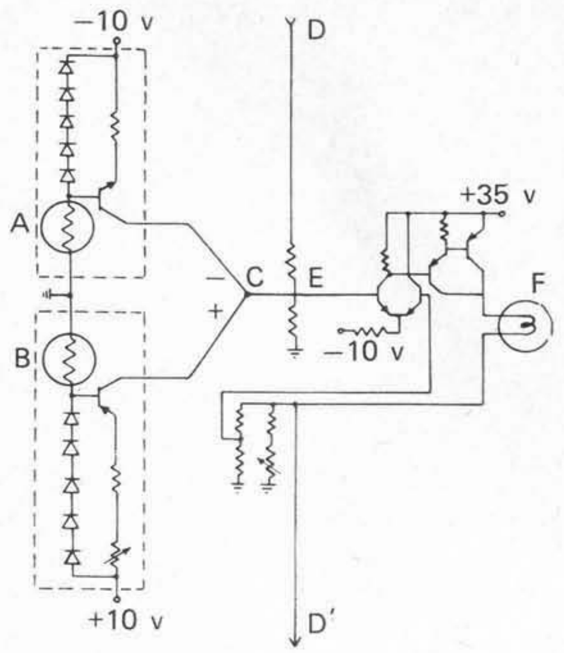


FIG. 11. Schematic diagram for one receptor pair of the electronic embodiment of the system. The output of the photocell A is logarithmically amplified and opposed to the logarithmically amplified output of the photocell B. The opposed signals are summed at C and then summed with the continued product D at E. In the machine, as contrasted with the scheme in Fig. 9(a), the signal is amplified to drive bulb F and to isolate the continued product output D' which is passed on to the next receptor pair. The bulb F in the schematic diagram is in the display panel. Because the bulbs were chosen so that, under the particular condition of their use, they have an antilogarithmic response, there is no separate antilog amplifier. The flux from the bulb corresponds with the lightness value computed by the sequential product up to that point on the chain.

APPROACHES TO A NEURAL SYSTEM

Although, in the machine, the locations of the edges are known, in the actual retinal image, the locations of the edges are of course unknown, yet it is important that a pair of receptors detect every edge. If any edge is not detected, all the subsequent readings along the path would be incorrect. If a great many pairs of receptors are closely packed, then the probability of missing an edge is small. Closest packing occurs in a system in which a single receptor is in the leading side of one bridge pair and the trailing side of the next bridge [Fig. 9(b)].

If we are considering a small number of receptor pairs, widely spaced (as shown in Fig. 7), it is sufficient to say that for each bridge the ratio of energies within an area of uniform reflectance approached 1.0; thus we eliminate the effect of uneven illumination. If, however, pursuing the purpose of capturing every edge, we consider very large numbers of tightly packed receptors, we must be concerned with the fact that across the expanse of an area of uniform reflectance, even though the ratios individually approach 1.0, the small deviation from each 1.0 for each pair leads to the accumulation of a substantial deviation from 1.0 on the exit side of the area. However, we suggest that in the physiologic model the very small systematic differences from 1.0 are lost at the synaptic thresholds, thus making feasible the continuum of contiguous cells in a chain of bridges. Such a chain cannot miss an edge and will not react to substantial but gradual changes of illumination across the field. Such a chain will also be completely indifferent to change of illumination as a function of time.

The schemes in Figs. 9(a) and (b) lead to a third important system. This is a biological system which derives the sequential product by means of excitatory and inhibitory synapses. In this biological system, the electrical concept of resetting is replaced by the hypothesis that an inhibitory signal greater than the signal coming down the chain will block it entirely rather than making it negative.[24] In this third scheme, a signal travels along the chain, being increased or decreased by successive receptors until it meets an inhibitory signal that is larger than the signal from the sequential product. At this point the old sequential product is blocked and the receptor K transmits its signal to its synapses I and E. Synapse I is an inhibitory synapse and operates on the signal from the sum J plus the sequential product N. This operation forms a new sequential product N' after I and is read out from the chain between the two synapses I and E. Synapse E is excitatory and combines its contribution with this new sequential product N', and so on.

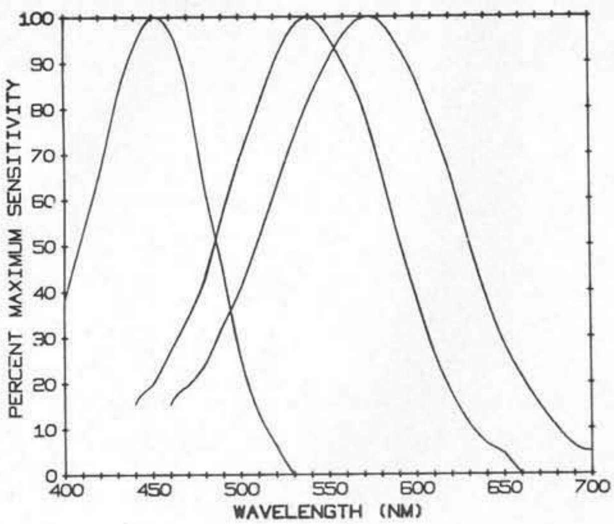


FIG. 12. Absorption curves of visual pigments. These curves we calculated using the Dartnall nomogram[28] and the maximum wavelengths 570, 535, and 445 nm.

In all the diagrams in Fig. 9, each receptor has been a member of only a single path. Just as a single receptor can be a member of two or more bridges on the same path, a single receptor can also be a member of many bridges on a number of different paths. The readout that corresponds to the reflectance of a point in an object can be the average of the sequential products read out for a number of different paths that pass through that point. The receptor can serve many paths, and the readout can average the sequential products of many paths at one point, without destroying the individuality and independence of each path.

By designing computer programs based on the lightness model we have described here, and by designing programs that relate computer to observers,[25] we have been able to arrive at a correlation between predicted lightnesses and observations. Because much of this computer work was carried on after the date of this lecture, we shall describe it in a later paper.

SEQUENTIAL PRODUCTS AND THE COLORED MONDRIAN

We have shown how a single retinex determines the lightnesses on the black, white, and gray Mondrian. The essence of retinex theory is that each retinex is served by one set of cones, that there are three retinexes and three sets of cones, and that each set of cones contains a pigment with an absorption curve different from that of the other two. All the cones with the same spectral sensitivity cooperate to generate the sequential product for their portion of the spectrum and for their retinex. In building the theory of the sequential product, we used the black, white, and gray Mondrian rather than the colored Mondrian. When we now turn to the colored Mondrian, we must take into account the facts that the three pigments in the three sets of cones have broad absorption curves, and that each colored area implies a variation of reflectance over the bandwidth of each set of cones. Because the three pigments in the three sets of cones have broad absorption curves[26–28] (Fig. 12) the signal fed into the neural pathway by a cone in one of these sets will depend on the integral of the spectral product of the absorbance of the pigment times the irradiance times the reflectance of the paper in the colored Mondrian. For these colored papers, the sequential ratio across an edge between rectangles will equal the ratio of the integrals, rather than the ratio of the luminances. (For the black, white, and gray papers, the sequential product obtained from a ratio of integrals is equal to that obtained from the ratio of the luminances.) It is important to emphasize that the integrals that form the sequential ratios for a given retinex use in the integrand the absorbance of the pigment for that retinex. The substitution of the ratio of integrals for the ratio of reflectances does not change the first important characteristic of the sequential product, namely, invariance with change of the over-all illumination of the whole Mondrian.

However, the variation of reflectance of a colored paper across the response band of one pigment is a weighting factor for a change of relative irradiance with wavelength for the flux from that paper. Therefore, for each retinex, the lightness scale for each of the colored papers in the Mondrian may shift a little with large changes of the wavelength composition of the illumination used by that retinex.

The experiment described at the beginning of this paper showed that changes of illuminance as great as the ratio of the reflectances do not change the color names of the papers. There were, however, some small changes and we can now understand these. If we compare the curves of the sharp-cut bandpass filters (Figs. 1 and 12) with the curves for the absorbance of the visual pigments, we see that each pigment, although absorbing principally from the light designed to illuminate it, also absorbs some light from the other filters. Therefore, for any one retinex, the values of the integrals at the junctions of the rectangles will vary somewhat with changes of the relative illumination from the three projectors. Consequently, there will be a small change of relative lightness of a given area on the three retinexes, a change small enough to be apparent on only a few of the colored rectangles.

The integral, over the absorption band of a pigment, of the product of absorbance times irradiance times reflectance, corresponds with the familiar integral in colorimetry, but is used here in an entirely different way. Here, the integrals are related to each other only for one retinex (by the sequential product); the three integrals for the three pigments for the flux coming from a point on a colored paper are not compared. Whereas the function of colorimetry is to classify reflectances into categories with similar visual properties, the function of retinex theory is to tell how the eye can ascertain reflectance in a field in which the illumination is unknowable and the reflectance is unknown.

ACKNOWLEDGMENTS

We wish to thank Julius J. Scarpetti for extensive experimentation in reproducing the Mondrian experiments and for the production of the transparency inserts. We are grateful to J. L. Benton and S. H. Perry for their assistance, comments, and criticism in all phases of this work, and to P. Dubois for his thoughtful discussions. In addition, we wish to thank M. A. Watson for her assistance with the many drafts of this paper.

REFERENCES

* Presented upon the award of the Frederic Ives Medal, 13 October 1967, Fall Meeting of the Optical Society of America [J. Opt. Soc. Am. **57**, 1428A (1967)]. For citation see J. Opt. Soc. Am. **58**, 597 (1968).

[1] Although sensations of lightness show a strong correlation with reflectances in most real-life situations, there are many important departures from this strong correlation. The color-contrast experiments of Chevreul[2] and Mach bands[3] are examples of such departures. In addition, in complex images, there are small but systematic changes of lightness when the over-all level of illumination changes (Jameson and Hurvich[4]; Bartleson and Breneman[5]). And, of course, any general theory must, as well, explain the simple situations in which surround comprises the entire environment (Hess and Pretori[6]; Wallach[7]; Stevens and Galanter[8]).

[2] M. E. Chevreul, *De la Loi du Contraste Simultane des Couleurs* (Pitois-Levrault, Paris, 1839).

[3] E. Mach, Sitzber. Math. Naturw. Kl. Kais. Akad. Wiss. **52**, 302 (1865).

[4] D. Jameson and L. M. Hurvich, Science **133**, 174 (1961).

[5] C. J. Bartleson and E. J. Breneman, J. Opt. Soc. Am. **57**, 953 (1967).

[6] C. Hess and H. Pretori, Arch. Ophthalmol. **40**, 1 (1884).

[7] H. Wallach, J. Exptl. Psychol. **38**, 310 (1948).

[8] S. S. Stevens and E. H. Galanter, J. Exptl. Psychol. **54**, 377 (1957).

[9] We avoided the use of a pattern of squares because previous experience had taught us the hazard of the superposition of afterimages as the eye moves.[10] Our completed display uses rectangles in an array the format of which reminded us of a painting by Piet Mondrian in the Tate Gallery in London. Thus we call our display the Mondrian.

[10] N. Daw, Nature **196**, 1143 (1962).

[11] E. H. Land, Am. Scientist **52**, 247 (1964).

[12] S. Hecht and Y. Hsia, J. Opt. Soc. Am. **35**, 261 (1945).

[13] E. H. Land, Proc. Natl. Acad. Sci. U. S. **45**, 115 (1959).

[14] E. H. Land, Proc. Natl. Acad. Sci. U. S. **45**, 636 (1959).

[15] E. H. Land, Sci. Am. **201**, 16 (May 1959).

[16] E. H. Land, Proc. Roy. Soc. (London) **39**, 1 (1962).

[17] J. J. McCann and J. Benton, J. Opt. Soc. Am. **59**, 103 (1969).

[18] Y. LeGrand, *Light, Colour and Vision, 2nd ed.* (Chapman and Hall, London, 1968), p. 225.

[19] Committee on Colorimetry, Optical Society of America, *The Science of Color* (Crowell, New York, 1953), p. 52 (available from Optical Society, Washington, D. C.).

[20] R. M. Evans, *An Introduction to Color* (Wiley, New York, 1948), p. 119.

[21] Reference 20, p. 159.

[22] Figure 6 was made as a transparency so that the photograph would be the best possible reproduction of the original experiment. The range of luminances of the original display was about 500 to 1. The reproduction must have a range of transmittances that approaches that range of luminances. In addition, the photograph must not alter the relative luminances of any areas by nonlinearities of the film response. It is very difficult to obtain both these properties in reflection prints, whereas the greater intrinsic dynamic range of a transparency allowed us to satisfy both conditions. In addition, the optical densities of each area across the horizontal midline of Fig. 6 are the same as those in Fig. 4.

[23] We are deeply indebted to L. Feranni and S. Kagan for developing the electronic representation of the system for finding the sequential product. The work on this display helped us to clarify our analysis.

[24] F. Ratliff, *Mach Bands: Quantitative Studies on Neural Networks in the Retina* (Holden-Day, San Francisco, 1965), p. 110.

[25] J. J. McCann, E. H. Land, and S. M. Tatnall, Am. J. Optom. Arch. Acad. Optom. **47**, 845 (1970).

[26] P. K. Brown and G. Wald, Science **144**, 45 (1964).

[27] W. B. Marks, W. H. Dobelle, and E. F. MacNichol, Science **143**, 1181 (1964).

[28] H. J. A. Dartnall, Bull. Brit. Med. Council **9**, 24 (1953).