

AMERICAN JOURNAL OF OPTOMETRY
and ARCHIVES of
 AMERICAN ACADEMY OF OPTOMETRY
 Established 1924

Vol. 47

November, 1970

No. 11

A TECHNIQUE FOR COMPARING HUMAN VISUAL RESPONSES
 WITH A MATHEMATICAL MODEL FOR LIGHTNESS*

John J. McCann†, Edwin H. Land‡, and Samuel M. V. Tatnall◆
 Polaroid Corporation, Cambridge, Massachusetts

ABSTRACT

This paper reports a technique for comparing the human visual responses with lightness predictions made by a mathematical model. The human visual responses are determined by having a number of observers compare the lightnesses in a Test Display with those in a Standard Display. The mathematical model's predictions are made by processing numbers that are identical to the luminances in the Test Display. These predictions are then scaled relative to the same Standard Lightness Display used by human observers. Methods of analyzing the results are discussed, as well as a variety of situations that can be used to establish whether a particular model can be considered a general model for lightness.

The purpose of this paper is to describe a technique for comparing the human lightness response with a mathematical model of that mechanism. Since lightness is in itself a complex, diverse problem, we feel that it is necessary to test any model for lightness in a variety of situations. Our method for comparing observers' results with any mathematical model's results for a variety of test situations will constitute the scope of this paper.

Lightness^{1, 2, 3, 4} is the family of sensations from white to black that a person sees. Lightness is the output of a biological system. It is a sensation. There is no physical definition for lightness because it is not necessarily related to a physical quantity of light from a point, either in radiometric terms or photometric terms.

Although it is commonly believed that there is a simple relationship between the amount of light coming from an object and how light or dark that object appears, there are many experiments that contradict that belief^{5, 6, 7}. As a particular example, let us study an experiment by Land. A display called a McCann⁸ Mondrian was made of various white, gray, and black papers. The papers were arranged so that the surround around each paper was arbitrary, multiple and variegated. The surround was arbitrary because there were no consistent patterns such as only low reflectance papers around high reflectance papers. The surround was multiple because there were many different papers around each

*Read before the Annual Meeting of the American Academy of Optometry, Philadelphia, Pennsylvania, December 16, 1969. For publication in the November, 1970, issue of the AMERICAN JOURNAL OF OPTOMETRY AND ARCHIVES OF AMERICAN ACADEMY OF OPTOMETRY.

†Scientist, B.A.

‡Director of Research, D.Sc.

◆Associate Scientist; B.A., A.M. Chairman, Science Department, William Penn Charter School, Philadelphia, Pennsylvania.

paper; it was variegated because the many different papers were significantly different from each other.

Land placed a lamp below and in front of the Mondrian display so that the illumination was non-uniform. Since the lamp was much nearer the bottom than the top, many more photons per unit area fell on the bottom of the display than the top. He then selected a high reflectance paper near the top and a low reflectance paper near the bottom. He adjusted the position of the lamp so that the same luminance was coming to the observer's eye from both areas. This was possible because the product of the higher reflectance and the lesser illuminance could be made equal to the product of the lower reflectance and the greater illuminance. The important point was that these two areas had exactly the same luminance, yet they did not look the same. The area at the top looked dramatically lighter than the area at the bottom. It is noteworthy that the lightnesses of these areas correlate strongly with reflectance.

We set out to find a mathematical model that could take the information at these two areas, as well as the information in the rest of the scene, and compute a lightness value that agrees with what we see for each area. We have already seen that the luminance of an area need not correlate with lightness. While there is generally a strong correlation between lightness and reflectance, there are phenomena such as Mach bands which show that this strong correlation does not always hold. Since the receptors in the retina respond to the luminous stimulus of objects, it would seem logical that the model should begin with luminance and then correct for departures from perfect correlation with lightness. We, however, took a different approach. Because in most situations there is a very strong correlation between lightness and reflectance, our model, although starting with luminance, will attempt to derive reflectances, and then make adjustments for imperfection in the correlation between lightness and reflectance. We therefore looked for a model that could determine the reflectance of any area under any condition of illumination without the usual constraints of photometry, such as placing the standard reflectance next to each area in the display. Having devised such a system, we began to modify that model to take into account the numerous situations in which lightness correlates less strongly with reflectance. The early work we performed was reported by Land in his Ives Medal Address, given to the Optical Society of America⁸. In that lecture Land described a mathematical model that could reproduce the lightnesses of a display independent of the luminances of each area. He also demonstrated a machine that embodied this mathematical model. Analogous to the Mondrian experiment, the machine produced two dramatically different outputs from some areas that were sending the same luminance to the machine. In addition, the machine gave the same outputs from areas which look alike to observers, yet were sending to the model different luminances.

Since the preceding experiments employed qualitative results, the approximation that the observers' sensations are directly related to reflectances was sufficient. If we are comparing several similar models that produce similar sets of predictions, we need more quantitative information to choose which model

is most like the human visual system. The obstacle preventing direct quantitative comparisons is that on the one hand the observer generates sensations, and on the other hand the model generates numbers. A technique has to be established that can relate sensations to numbers. We have seen that, although there is no simple relationship between lightness and luminance, there is a strong correlation between lightness and reflectance. This suggests that a particular situation might be found in which lightness and reflectance exactly correlate.

If we could establish a display showing a unique relationship between reflectance and lightness, we could use this display as a basis of comparing the observer's lightnesses with the model's numbers. The observer could describe any sensation as equivalent to a sensation found in such a Standard Lightness Display. Since each sensation in this Standard Display would have a unique reflectance associated with it, we could then assign an Equivalent Reflectance to the sensation we wish to quantify.

STANDARD LIGHTNESS DISPLAY

The problem becomes one of describing a display which generates sensations that are related to reflectance in a known, unique way. Imagine a display with the following three properties. First, it contains a set of many different Standard Reflectances. Second, the entire display is illuminated uniformly; that is, the same amount of light is falling on all parts of the display. Third, the same complex surround is around each Standard Reflectance; that is, whatever reflectance papers are around one Standard Paper must be around all Standard Papers (Fig. 1).

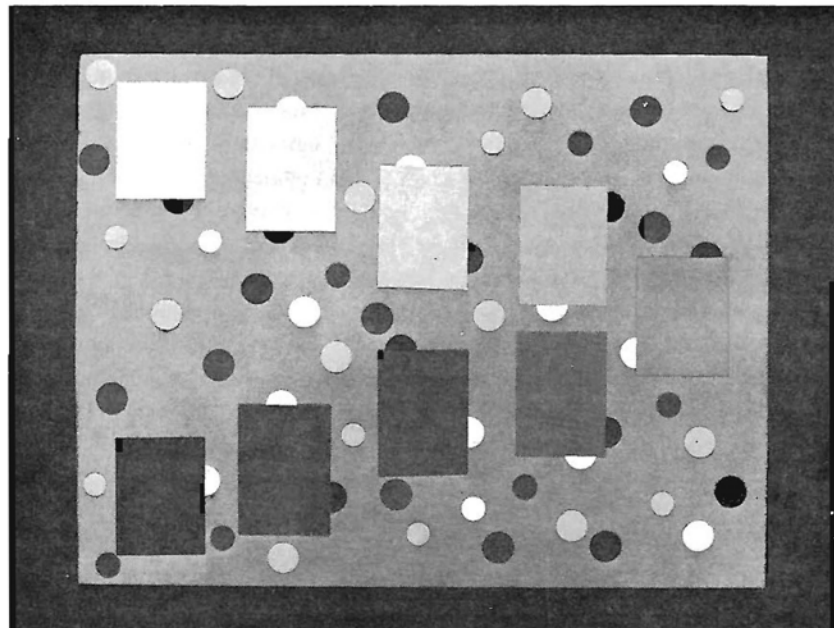


Fig. 1. Photograph of Standard Lightness Display.

The next step is to put these Standard Papers into two rank orders: The

first, the rank order of lightness from light to dark; the second, the rank order of reflectance from high to low. We have always found that under the conditions of uniform illumination and identical surrounds, the two rank order numbers for a particular paper will always be the same. Now imagine adding a new member to the set of Standard Papers; for example, the lightness of the new paper might be between the lightness of the third and fourth Standard Papers. We have always found that the reflectance of the new paper will lie between the reflectances of the same two papers—in this case between the reflectances of the third and fourth papers. This new paper is now a member of the set of Standard Papers, and we can look for the next new paper to insert in the set of Standard Papers. Our experience leads us to the hypothesis that every member we add between the lightnesses of two Standard Papers will have a reflectance between the reflectances of those two Standard Papers. If this hypothesis is true, then with uniform illumination and identical surround there is a unique number associated with any sensation from white to black. If we now consider a lightness sensation that is generated in any other display, in certain illuminations, and in any surround, we can find the matching sensation among the Standard Papers. If we assign the reflectance number of the Standard Paper to the area that generated the identical sensation, we will have assigned an Equivalent Reflectance to that sensation, thus quantifying it. We then assign the mathematical model the task of computing the Equivalent Reflectances from the display's luminances. It would be difficult to construct a Standard Lightness Display that would contain enough papers to generate all sensations from white to black. Therefore our actual Standard Display contains only nine papers which are used as guideposts for the observer to assign Equivalent Reflectances.

How should these Standard Papers be chosen? If they are equally spaced in reflectance from the highest to the lowest reflectance, they will not be equally spaced with respect to lightness. As we will describe later, our technique of analysis demands that our Standard Papers be equally spaced in lightness.

The problem is that there are two types of commonly used lightness scales: the Munsell Value scale and the Stevens Power Law scale. As Stevens and Galanter⁹ point out, the explanation for the different scales is the different experiments that generated them. Our only decision was to evaluate whether the experimental question we were asking our observers was derived from the Munsell difference technique or the Stevens ratio technique. Munsell¹⁰ specifically instructed his observers to construct an equally spaced lightness scale. Newhall, Nickerson and Judd¹¹ improved Munsell's scale and again asked the observer for equal lightness spacing. Further, Newhall¹² showed by a different technique that observers found the Munsell Value scale to be equally spaced. We therefore decided to use Munsell's procedure for choosing the Standard Papers.

We must now determine whether the Munsell scale is the same as an alagous scale established under our conditions of illumination and surround. We would like a surround that approximates, at least in some ways, the real world. A uniform black or white surround is not typical, because real objects are viewed in a more complex, variegated environment. We therefore made a sur-

round with small circular pieces of white, gray, and black papers distributed randomly on a gray paper. A semicircle of white and a semicircle of black was placed contiguous to each Standard Reflectance, as shown in Fig. 1.

Using essentially Munsell's procedure¹⁰, we asked observers to establish an equally spaced lightness scale in our illumination and with our surround. The 20 papers on the surround were illuminated uniformly by a bank of fluorescent tubes overhead. Ten observers were asked to select a piece of paper that looked halfway between the whitest and blackest papers. The observers were then asked to select papers that were midway between each of these three papers: white, the chosen middle-gray, and black. The observers then subdivided the scale again and selected four more pieces that were halfway between the five pieces already chosen. Thus they had selected a nine-step scale from white to black.

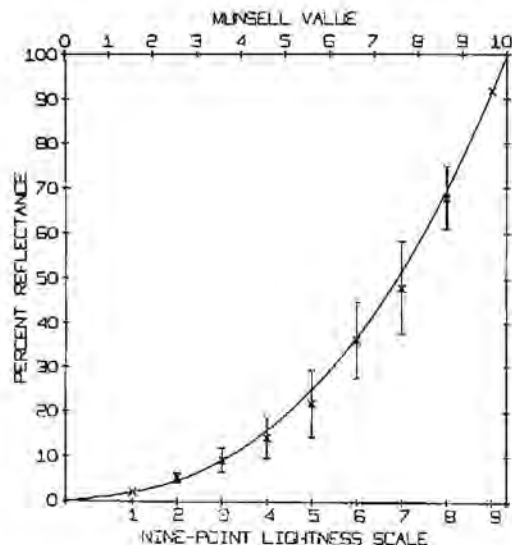


Fig. 2. Per cent absolute reflectance plotted as a function of lightness on an arbitrary nine-point scale (lower horizontal axis) and on the Munsell Value scale (upper horizontal axis). The X's are the means of reflectances chosen by ten observers for a nine-point equally spaced lightness scale. The solid curve is Newhall's empirical fifth-degree polynomial relating reflectance to Munsell Value modified by LeGrand to give absolute reflectance.

Fig. 2 is a graph of reflectance *vs.* lightness. On the bottom horizontal axis we have placed our nine-point lightness scale. On the top horizontal axis is placed the Munsell Value scale. The X's are the means of our ten observers' choices for equally spaced lightnesses. The vertical line through each X represents plus and minus the standard deviation of the observers' results. The solid curve is Newhall, Nickerson and Judd's¹¹ empirical polynomial relating reflectance (relative to MgO) to Munsell Value, modified as suggested by LeGrand⁸ to give absolute reflectance. The graph shows that there is no significant difference between our results and the polynomial of Newhall, *et al.* Thus we chose simply to use the shape of Newhall's polynomial rather than finding a similar curve that fit our data. We did not consider our experiment a substi-

tute for the extensive work done on the Munsell Value scale, but only a check to see whether our experimental conditions introduced any significant deviations. We decided to maintain our nine-point scale, which is a simple linear transformation of the Munsell scale. We think that the relatively large standard deviations are probably due to the fact that our observers had only a third as many papers to choose from as Munsell observers; if our observers had a greater variety of papers to choose from the standard deviations might be reduced considerably.

TABLE 1
PER CENT ABSOLUTE REFLECTANCE

Step	Mean of Observers' Choices	Predicted by Newhall's Curve	Actual Used
9	92	92	92
8	68	70	70
7	48	52	51
6	36	37	36
5	22	25	27
4	14	16	17
3	9.2	9.1	11
2	5.2	4.7	6.0
1	2.0	2.0	2.0

Table 1 shows the means of our experimental results, reflectances computed from the Newhall curve, and the reflectances of pieces of paper that were used to make the completed Standard Lightness Display shown in Fig. 1. We used this display as follows: The observer quantifies the lightness of a particular area by comparing it to the various steps on the Standard Display. If an area looks identical to step 5 in the Standard Display, that area has lightness number 5. If, however, the area appears in lightness to be between steps 5 and 6, the observers are asked to interpolate to the first decimal place.

This display also establishes the values of the predictions made by the mathematical model. As we have already mentioned, the model must predict the value of the Equivalent Reflectance of each area. If we have an equation for lightness numbers as a function of reflectance, we can convert the model's Equivalent Reflectance to lightness numbers that are scaled identically with the observers' lightness values. Newhall's equation, plotted in Fig. 2, is a fifth-degree polynomial giving reflectance as a function of Munsell Value. We need a linear transformation of the inverse of Newhall's polynomial, that is, lightness on our nine-point scale as a function of reflectance. Rather than trying to find the inverse of Newhall's function, we fitted the curve

$$V = 1.396 R^{0.433} - 0.147$$

where V is Munsell Value and R is absolute reflectance, to the tabulated values given in LeGrand. With the linear transformation to our nine-point scale, the equation takes the form

$$L = 1.371 R^{0.433} - 0.645$$

where L is lightness on our scale.

COMPARISON TECHNIQUE

Having discussed a Display that is our metric, we need to discuss the tech-

nique of comparing any area with this Standard Display. One technique would be to ask the observers to memorize the appearance of each paper in the Standard Display. Then in further experiments the observer would compare the sensations produced by a Test Display with his memory of the Standard Display. This technique has the advantage that the Standard Display cannot influence the Test Display. The problem with this procedure is that it introduces a sizable uncertainty into the results. If the observers' uncertainty is large, we would be unnecessarily lax in testing the accuracy of our mathematical model.

A better technique would be to make direct comparisons of the Test Display and the Standard Display at the same time. The observers could look at the Test Display with one eye and the Standard Display with the other eye. This technique has greater intuitive appeal because the observer is looking at both Displays simultaneously, but, since they are seen with different eyes, one Display presumably does not affect the other. From a more practical point of view, the simultaneous comparison technique has the following problems: The observer usually has to view the targets with optical aids: he will see retinal rivalry if the images overlap, and he is using different parts of the retina if they do not. The observer often finds that the experiment is difficult and very fatiguing. Considering the number of separate experiments we planned to conduct, we felt that this technique would be too taxing on the observers.

These disadvantages of simultaneous comparison are eliminated by closely spaced sequential comparisons. Here the observer looks at the Standard Display through a window. The Standard Display is the only thing in his field of view. He can study the Standard Display as long as he desires. He then moves his head to look at the Test Display which now is the only thing in his field of view. He can move his head back and forth as many times as is necessary for him to decide the areas that most closely match in the Standard and Test Displays.

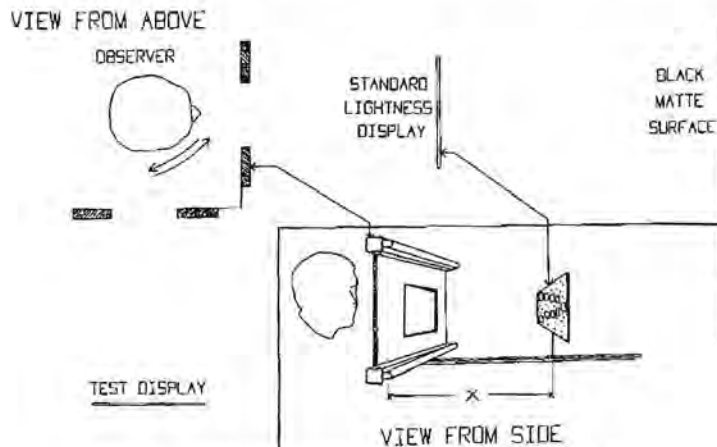


Fig. 3. View of apparatus from above and from side.

Fig. 3 is a view of the apparatus from above. The observer looks through one of the windows in a partition in front of one of the displays. The observer

need only turn his head 90° to see the other display that is mounted similarly. Since the observer can make comparisons of different areas in the other display by moving his head, this arrangement has the convenience of comparing two widely separated objects in the same room, yet fulfills the demand that each display be viewed separately. The border that the observer sees around the display is a wall, covered with black matte-surface paper, six feet behind the display. Thus a very small amount of light is coming from the outside black surround.

Fig. 3 shows a side view of the apparatus. A 9 by $11\frac{1}{2}$ -inch display is mounted on a thin black steel shaft that stands in any of a series of holes in the base. By making the distance X between the planes of the display and the lights large and by using two fluorescent tubes, we can obtain uniform illumination. If we desire non-uniform wedge illumination, we can turn off one of the fluorescent tubes and control the slope of the wedge by varying the distance X .

The level of over-all illumination is carefully controlled. In uniform illumination, the Test Display receives the same illuminance as the Standard Lightness Display. In non-uniform illumination, the area that receives the greatest illuminance receives the same illuminance as the Standard Lightness Display.

ANALYSIS OF RESULTS

We now discuss the way in which we compare the lightness numbers generated by a mathematical model with those reported by our human observers. One possible technique for comparison would be to determine a criterion for a correct prediction and then determine the proportion of correct predictions. A reasonable criterion would be that the prediction must fall within a particular range of results determined by the mean plus and minus one standard deviation of all the observers' results. Some models may report many lightness numbers within an area for which an observer gives only one report. We can also compute the mean and standard deviation of the model's predictions for such areas. Then the mean of the model's predictions, plus and minus one standard deviation, must intersect the mean human observation, plus and minus one standard deviation, to be considered correct.

A better analysis, however, would represent the values of these statistics so that the sizes of the standard deviations and coincidence of the means would not be lost. An effective presentation would be a plot of the mean of the model's predictions against the mean of the observers' results for each area. Since both the model's numbers and the observers' lightnesses are scaled relative to the Standard Lightness Display, both axes of this graph are identical. Therefore, if the mean prediction is identical to the mean observation for each area, the plot of these points would describe a line at 45° . If we now include standard deviations as well as means, we have a set of boxes that should fall on the 45° line if the predictions are correct.

The demand for equally spaced lightness steps in the Standard Lightness Display comes directly from this inspection technique. For one to be able simply to look at this graph and evaluate the results, equal distances on the

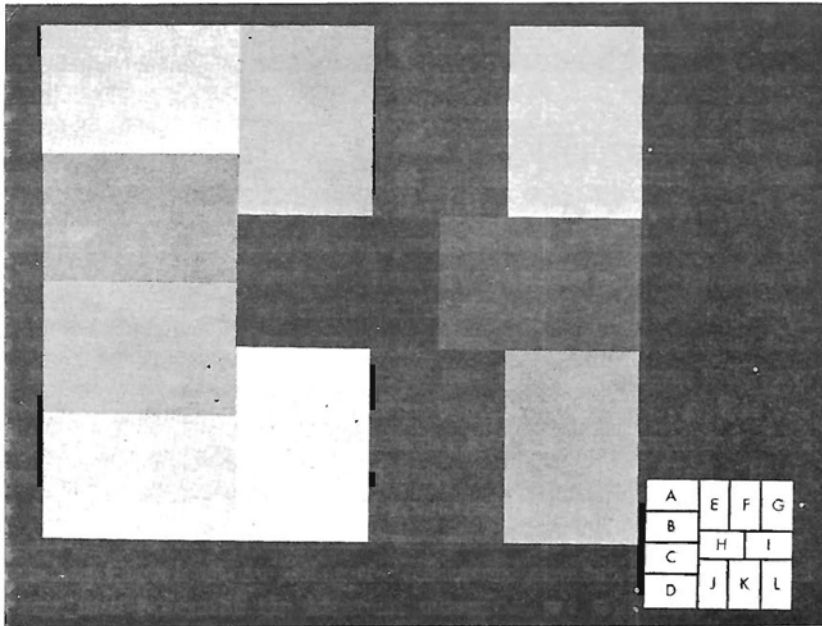


Fig. 4. Photograph of Tatami Test Display. Inset at right labels each area.

graph must represent equal intervals in lightness. If, instead, we used Equivalent Reflectances as axes, an error in the model's results near white would appear small to the observer, whereas the same size error near black would appear large.

Fig. 4 shows a simplified Mondrian which we call a Tatami. The twelve different papers are arranged to simulate a Japanese mat, hence its name. This Tatami has papers from all parts of the lightness scale. Although it is a greatly simplified representation of a normal scene, it fulfills our conditions that the surround of each area be arbitrary, multiple, and variegated. The Tatami was placed in our apparatus and was illuminated non-uniformly. The plane of the display was five inches from the axis of the lower fluorescent tube, which was the only one used. Measurement of two separate areas having the same reflectance (A and J in Fig. 4) shows that there is eight times more illuminance at the bottom than at the top.

Fig. 5 shows the graphical analysis of a mathematical model for lightness incorporating the principles described by Land⁸. It is beyond the scope of this paper to describe the many details of this model; nevertheless, this graph is presented as an example of our experimental results. The twelve areas in this Tatami have been labeled A, B, C, . . . J, K, L, as shown in the inset to Fig. 4. Each box in Fig. 5 is labeled with the letter of the area which it represents. The center of each box is the mean of the model's predictions and the mean of the observers' responses. The vertical side of each box is the mean model prediction, plus and minus one standard deviation; the horizontal side is the mean observer response, plus and minus one standard deviation. If the model's means and standard deviations happened to be identical to the observer's means and standard deviations the twelve rectangles in Fig. 5 would be squares

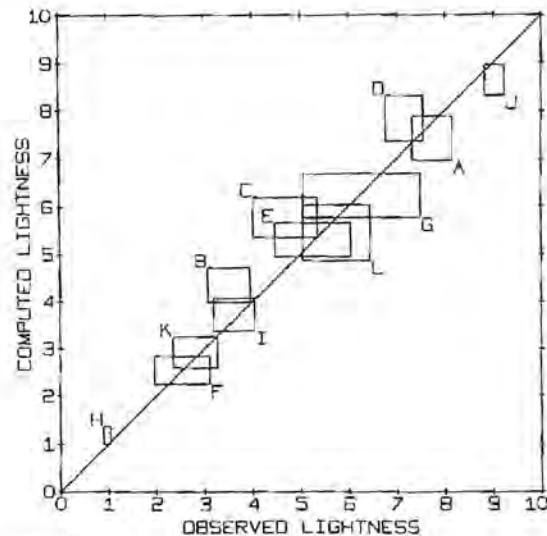


Fig. 5. Tatami Test Display lightnesses predicted by a mathematical model plotted against lightnesses reported by 10 human observers. The center of each box is the mean model prediction and the mean observer result. The sides of each box are the mean plus and minus one standard deviation of each result.

and their centers would lie on the line at 45° . For this particular experiment, the model agrees very closely with the observers for all regions of the lightness scale. Thus the technique of analysis is simply that of inspecting how well the boxes fall on the line drawn at 45° . Although we have investigated a variety of statistical schemes for analyzing whether one set of results is better than another, we have always returned to this simple graphical presentation because it retains most of the information about each experimental result.

DISCUSSION

The final topic to discuss is the variety of Test Displays that might be used as a general test of a particular model. A second Tatami could be made of predominantly high-reflectance papers, while a third could be made of predominantly low-reflectance papers. These Tatamis are important because they would have average luminances that differ from that of the normal Tatami. They provide a good point for comparison of our model with other mathematical models for predicting lightnesses which, unlike our own, compute and use average luminances. These three Tatamis could be used in testing the model's ability to simulate the visual system's responses to both uniform and non-uniform illumination.

The classical situations of a gray piece of paper on a white surround and that same paper on a black surround are important Test Displays. These Displays are probably the most recognized examples of the departure from correlation of lightness with both reflectance and luminance.

We also feel it is important to include in a general test of a lightness model some more unusual but theoretically important situations. Mach bands and the edge phenomena described by Cornsweet (in Ratliff¹⁴) are such phenomena.

Mach bands can be generated by a step gradient made of pieces of uniform reflectance papers arranged in order. One side of each paper is noticeably lighter than the other, despite the fact that the luminance across each paper is constant. Similarly, at both ends of a continuous gradient either a light or dark Mach band appears, despite the fact that there is no luminance or reflectance that can correlate with these sensations. Cornsweet's experiment shows that two entire areas of the same reflectance can be made to look different in uniform illumination by an edge composed of a low-slope gradient, a sharp edge, and a second low-slope gradient complementary to the first.

Although these test situations do not necessarily provide a complete general test of any model, a particular model that can predict lightnesses in all these situations is certainly of great value. We propose to study all of these situations using the Standard Lightness Display described here to quantify both observers' sensations and model's predictions on the same equally spaced scale. Thus, using the techniques described in this paper, we propose to test our present ideas, shape our future models, and find, if possible, a single general model that can predict the lightnesses in all the different situations discussed.

ACKNOWLEDGMENT

We thank Marie A. Watson and Thomas H. Taylor for their invaluable help in preparing this paper.

REFERENCES

1. Evans, Ralph M., *An Introduction to Color*. New York, John Wiley & Sons, 1948, p. 119.
2. Optical Society of America. Committee on Colorimetry, *The Science of Color*. New York, Thomas Y. Crowell Co., 1953, p. 373.
3. LeGrand, Yves, *Light, Colour, and Vision* . . . translation by R. W. G. Hunt and others. 2d ed., New York, John Wiley & Sons, 1968, pp. 226-228.
4. Land, Edwin H., *The retinex*. *Am. Scientist*, 52: 247-264, 1964.
5. Chevreul, M. E., *The Principles of Harmony and Contrast of Colors and Their Applications to the Arts*, translated by Charles Martel, 1854. Reprinted, New York, Reinhold Corporation, 1967.
6. Albers, Josef, *Interaction of Color*. New Haven, Yale University Press, 1963, Chapters 4-6.
7. Mach, Ernst, On the effect of the spatial distribution of the light stimulus on the retina (translation), in *Mach Bands: Quantitative Studies on Neural Networks in the Retina*, by Floyd Ratliff. San Francisco, Holden-Day, Inc., 1965, pp. 253-271.
8. Land, E. H., and J. J. McCann, *Lightness Theory*. *J. Opt. Soc. Am.* 61:1-11, 1971.
9. Stevens, S. S., and E. H. Gallanter, Ratio scales and category scales for a dozen perceptual continua, *J. Exper. Psychol.*, 54: 377-411, 1957.
10. Munsell, A. E. O., L. L. Sloan, and I. H. Godlove, Neural value scales. I. Munsell neural value scale, *J. Opt. Soc. Am.*, 23: 394-411, 1933.
11. Newhall, S. M., Preliminary report of the Optical Society of America Subcommittee on the Spacing of the Munsell Colors, *J. Opt. Soc. Am.*, 30: 617-645, 1940.
12. Newhall, S. M., D. Nickerson, and D. B. Judd, Final report of the Optical Society of America Subcommittee on the Spacing of the Munsell Colors, *J. Opt. Soc. Am.*, 33: 385-418, 1943.
13. Newhall, S. M., A method of evaluating the spacing of visual scales, *Am. J. Psychol.*, 63: 221-228, 1950.
14. Ratliff, Floyd, *Mach Bands: Quantitative Studies on Neural Networks in the Retina*. San Francisco, Holden-Day, Inc., 1965, p. 74.