

# Lightness, equivalent backgrounds, and anchoring

NICOLA BRUNO and PAOLO BERNARDIS  
*University of Trieste, Trieste, Italy*

and

JAMES SCHIRILLO  
*Wake Forest University, Winston-Salem, North Carolina*

Observers compared two center/surround configurations haploscopically. One configuration consisted of a standard surface surrounded by two, three, or four surfaces, each with a different luminance. The other configuration consisted of a comparison surface surrounded by a single annulus that varied in luminance. Center surfaces always had the same luminance but only appeared to have the same lightness with certain annuli (equivalent backgrounds). For most displays, the luminance needed to obtain an equivalent background was close to the highest luminance in the standard surround configuration. Models based on the space-average luminance or the space-average contrast of the standard surround configuration yielded poorer fits. Implications for computational models of lightness and for candidate solutions to the anchoring problem are discussed.

Consider two surfaces cut from the same gray paper. One is pasted in the center of a background that consists of several other adjoining papers—some are black, some are white, and some are various shades of gray. The other is pasted in the center of a single piece of paper. Which shade should be chosen for this single background to cause the perception of the two center papers to appear equal? According to many people's intuition, almost any shade would do. After all, the two center surfaces are cut from the same material, and materials tend to preserve their color as they are moved in front of different surrounds. Nonetheless, this type of lightness constancy is not perfect. For instance, the same surface may become slightly lighter in front of a dark background and slightly darker in front of a light background. Thus, one would expect that at least some shades would *not* work. But which ones?

This paper presents three experiments aimed at answering this question. In an investigation of achromatic surface color (lightness), the question is unusual. Investigators have traditionally been interested in the constancy of surface lightness due to changes in the illumination (see, e.g., Arend & Goldstein, 1987; Gilchrist, 1988; Katz, 1911; Wallach, 1948). In the present experiments, we studied the change of lightness as a function of changes in the composition of the surround while the luminance of the assessed surface was kept constant. In-

vestigators have also investigated lightness constancy under proximal changes in luminance ratios—for instance, when the same surface is moved from one annular surround to another (see, e.g., Arend & Spehar, 1993b). In the present experiments, we surrounded a center surface with several patches of constant average luminance, so that the proximal ratio of the center to the average surround luminance remained constant. The number of patches in this surround was varied, and, for each different surround, the question was asked whether a single shade of gray can be found that will cause the two center surfaces to have the same lightness. These single shades of gray are called *equivalent backgrounds*, in that they appear to have an equivalent contrast effect to that of the multiple-patch surrounds. Our interest in equivalent backgrounds stems from their relation to one of the fundamental problems in lightness perception, the anchoring problem.

## ANCHORING

Computing achromatic surface color requires two processes: the spatial integration of the luminance ratios that are formed at each surface boundary within a scene, and the selection of a common reference. The aim of the first process is to define a scale of relative gray values (Arend, 1985; Blake, 1985; Gilchrist, Delman, & Jacobsen, 1983; Horn, 1974; Hurlbert, 1986; Land & McCann, 1971; Whittle & Challands, 1969). The aim of the second process is to "anchor" such a scale to a common reference, so that each surface can be assigned an absolute lightness (Cataliotti & Gilchrist, 1995). To date, three means of computing the reference term have been proposed: space-average luminance (Helson, 1943, 1964; Hurlbert, 1986; Hurlbert & Poggio, 1988; Judd, 1940; Land, 1986; Land, Hubel, Livingstone, Perry, & Burns, 1983), space-average contrast (Brown & MacLeod, 1991), and the highest lu-

---

This work was supported in part by CNR 9504106.11. The present research was discussed at the Trieste Encounter in Cognitive Science: "Illumination in Object-Oriented Vision," SISSA, Trieste, Italy, June 1995. The authors are indebted to the participants and to Robert O'Shea, Alan Gilchrist, and two anonymous reviewers for comments on the first paper submission. Correspondence should be addressed to N. Bruno, Dipartimento di Psicologia, Università di Trieste, via dell'Università, 7, 34123 Trieste, Italy (e-mail: nicola.bruno@univ.trieste.it).

minance of a surface within a scene (Land, 1977; Land & McCann, 1971; Wallach, 1948).

A number of recent results suggest that, in many cases, the human visual system behaves as if it uses the highest luminance in the scene as the reference. For instance, if the inside of a large dome is painted half black and half middle gray, the middle gray half appears almost white and the black half appears middle gray (Li & Gilchrist, 1993). If the space-average luminance is held constant in a Mondrian (i.e., a collection of gray patches of various lightnesses) while the highest luminance is varied, surface colors change noticeably. Whereas, if the highest luminance is kept constant while the average is varied, colors change only slightly (McCann, 1989, 1992). Finally, the main rationale behind averaging, the "gray world" assumption (Hurlbert, 1986), has been challenged by statistical analyses of naturally occurring reflectance spectra (Brown, 1994).

Although the weight of the evidence is shifting in favor of the highest luminance reference, arguments in favor of space-averaging remain. One argument is certainly that the solution to the anchoring problem is simplified when tied with known mechanisms in the visual system. For instance, Ganzfeld results are consistent with the notion that space-average luminance may be computed using a low-pass channel to estimate the overall intensity in a scene (Knau & Spillmann, 1994; Schubert & Gilchrist, 1992). Normalization based on space-average contrast may be implemented by the mechanisms hypothesized to be responsible for contrast gain control (Geisler & Albrecht, 1992; Robson, 1991).

In our present work, however, we were not directly concerned with issues of underlying mechanisms. Instead, we were primarily concerned with the implications of the equivalent background question for candidate solutions to the anchoring problem. Suppose, for instance, that a surface is surrounded by many colors, one of which is white. The highest luminance hypothesis predicts that surface lightness is based on the ratio to the white anchor. Therefore, the surface should have the same lightness as a surface of equal luminance surrounded by this white alone. Suppose, conversely, that the white is the surrounded surface itself. In this case, it should not matter which single surface is put around it. Any shade of gray will allow this white to appear the same as another white surrounded by several patches. Given these strong predictions of the highest luminance hypothesis, and given that these predictions differ dramatically from those based on the other candidate anchoring schemes, it is important to determine whether the prediction holds and, if not, it is important to understand the conditions under which it fails.

### EQUIVALENT BACKGROUNDS

Crawford (1947) provided the basis for the original formulation of an equivalent background principle. He compared the time course of the recovery of sensitivity after exposure to flashes having a fixed luminance

and variable size with the sensitivity to steady fields of a fixed size but variable luminance. The relationship between the time-course curves and the threshold-versus-intensity curves suggests that the course of dark adaptation can be approximately equated by the presentation of progressively weaker background lights. This may be interpreted as evidence that the global adaptive state of the retina is also represented more centrally within the system. This representation would be essentially equivalent to a form of spatial averaging, if one assumes that the global adaptive state depends on average retinal illuminance.

More direct evidence for spatial averaging was recently provided by an observation reported by Valberg and Lange-Malecki (1990). They placed an achromatic surface in the middle of a surround containing many other colors and then compared it with another central surface surrounded by a single adjustable gray background. The two central surfaces had the same reflectance, but they reportedly appeared to have the same lightness only when the single background was a space-weighted average of the complex surround. Depending on the nature of the weighting, therefore, this result may be taken as evidence for either the space-average luminance hypothesis or the space-average contrast hypothesis.

Other equivalent background studies, however, suggest that, in certain conditions, observed equivalent backgrounds are quite different from those predicted on the basis of spatial averaging. Schirillo and Shevell (1993) presented a checkerboard surrounding a standard center surface. A four-patch checkerboard had two different luminances that could be varied in contrast from 0.2 to 1 while keeping the space average constant. Observers adjusted the brightness of this center surface to match another surface surrounded by a homogeneous annulus having a luminance equal to the space average of the checkerboard. If the equivalent background is the space-average luminance, matches should have remained constant as checkerboard contrast varied. Their results, however, were only consistent with the space-average hypothesis for decrements relative to the homogeneous surround, while increments appeared increasingly dimmer as checkerboard contrast increased.

Additional studies are also at odds with the space-average luminance hypothesis. Bruno (1992) presented Mondrians consisting of six patches. The luminance of these patches spanned about 2 log units, and they could have three different sizes. Centered on each Mondrian was a standard surface. Observers compared this surface with a comparison surface of equal luminance but surrounded by a homogeneous annulus. The images were presented sequentially, with an intermittent blank field set to the space-average luminance. Observers adjusted the luminance of the homogeneous surround until the test and comparison center surfaces appeared to have the same lightness. The results exhibited two main features. First, observers often reported that no adequate surround could be found. Subsequent analyses indicated that these reports tended to occur when the standard patch had an in-

intermediate luminance relative to the surround luminances of the Mondrian. Second, when adequate surrounds could be obtained, they tended to be close to those predicted by the highest luminance rule.

## OVERVIEW OF THE EXPERIMENTS

The goal of the present experiments was to resolve these conflicting results while determining which hypothesis best predicts how the reference term is computed by human observers. The methodology used was a derivation from the studies of Bruno (1992) and Schirillo and Shevell (1993), with the addition of two new features. First, the displays were viewed haploscopically. In comparison with free binocular viewing, this feature has the advantage of controlling unwanted spatial interactions between the experimental surrounds. Second, observers performed either a *lighter-darker* or a *same-different* comparison for several pairs of configurations rather than adjusting either the center surface or the surround. In comparison with the method of adjustments, this feature has the advantage of minimizing temporal interactions between the presented stimuli. There are other ways to control for such unwanted interactions, such as procedures where observers are presented the displays for experimentally controlled intervals and make the adjustments during intertrial periods. Such methods, however, require remembering the presented percept for a short interval and may therefore introduce another source of variation in the matches. In addition, our simultaneous comparison method provides a metric of similarity between the pairs as well as a value corresponding to a point of subjective equality. Data about the degree of similarity between the judged surfaces at different values of their surrounds may provide important information that is not conveyed by a single adjustment value.

In each experiment, these two kinds of comparisons were performed by independent groups of participants. In one group, observers reported whether the left center surface or the right center surface appeared darker. In the other group, they reported whether the two center surfaces appeared either the same or different. Within each pair, the center surfaces always had the same luminance. The standard center was surrounded by two, three, or four patches. The comparison center was surrounded by a single annulus that varied in luminance across trials.

Equivalent background values were obtained by linear interpolation<sup>1</sup> from the psychometric curves of the *lighter* judgments. If there is a single value that functions as an equivalent background, a point of subjective equality for the two central squares exists when these curves cross the 50% line. This point should also correspond to the comparison surround luminances yielding the highest percentage of *same* judgments. If no value functions as an equivalent background, or if more than one does, then the psychometric curves either should never cross 50% or should cross 50% at more than one point. At the same time, the distributions of *same* judgments should have more than one peak.

## GENERAL METHOD

### Observers

For each of the three experiments, 60 members of the University of Trieste community volunteered. Thirty performed the *same-different* task, and 30 performed the *lighter-darker* task.

### Equipment

All stimuli were generated using a Silicon Graphics Indigo workstation and displayed on a carefully calibrated Silicon Graphics monitor. This monitor has a resolution of  $1,280 \times 1,024$  pixels and 256 simultaneously displayable gray levels covering a range of approximately 2 log units of luminance. Independent control of the images to each eye was achieved with a Crystal Eyes field-sequential system. In this system, an emitter broadcasts an infrared signal to switch liquid crystal lenses on and off in exact synchronization with the monitor at a rate of 60 Hz per eye. The lenses are mounted on glasses worn when observing the displays.

Monitor calibration was performed in two steps. First, photometer readings were obtained for the darkest and brightest grays that could be produced on the monitor, and the contrast and brightness switches were adjusted to achieve a range of about 2 log units. Next, luminances at different monitor locations and at different gray values were measured, and appropriate polynomials were fitted to obtain an equation for converting software-specified intensities to desired luminances. The spatial inhomogeneity in monitor output was minimal in the center area used for displaying the stimuli. The observers entered all of their responses using two mouse buttons.

### Haploscopic Presentation

The method of presentation is illustrated in Figure 1. In this method, different images are presented separately to each eye to prevent any retinal interaction between them. However, when the images are fused, observers perceive them as a single, composite image. In all three experiments, both images consisted of an overall background field enclosing a smaller configuration. The standard configuration was presented on the left side and consisted of a center square surrounded by two, three, or four patches (two patches shown) of the same space-average luminance. The comparison configuration was presented on the right side and consisted of a center square surrounded by a single annulus. Within each trial, the lu-

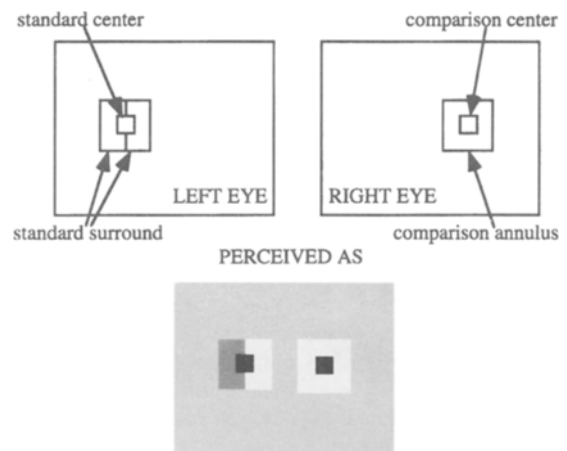


Figure 1. Schematic of the haploscopic presentation and terminology for the experimental displays of Experiment 1. Experiments 2 and 3 used the same method of presentation and the same configurations, except for the standard surround, which had either three or four surfaces of different luminances, instead of two.

minance of the two central surfaces was the same. The luminance of the comparison annulus was varied between trials across a range of 10 to 90  $\text{cd}/\text{m}^2$ , in steps of 10  $\text{cd}/\text{m}^2$ . To avoid calibration problems at very low monitor intensities, the overall background field was always set to the space average of all displayed luminances, except those of the single comparison annulus. As a consequence, the overall background luminance was constant for all pairs of a given center surface luminance, but it varied slightly across groups of pairs having different center surface luminances. Being set at the average, however, this overall background did not change the space-average value of a given display, nor did it affect the highest luminance. For these reasons, its potential effects are negligible. The haploscopic image was observed from a distance of 0.5 m in an otherwise dark room, and it subtended  $40^\circ$  horizontally and  $30^\circ$  vertically. The central squares and their surrounds subtended  $3^\circ \times 3^\circ$  and  $12^\circ \times 12^\circ$ , respectively.

### Two-Alternative Forced Choice (2AFC)

In all experiments, different observers performed one of two 2AFC tasks. One group indicated whether the two central squares appeared to depict surfaces cut from the same gray paper. Another group indicated which of the two central squares appeared to depict a lighter gray shade.

## EXPERIMENT 1

In Experiment 1, equivalent background values for bipartite surrounds were determined. These surrounds were comparable to those used by Schirillo and Shevell (1993) in that they consisted of only two different luminances. A separate pilot study replicated Schirillo and Shevell's reported effects of contrast between the luminances in the bipartite surrounds and of the central surface luminance. However, because the aim of Experiments 2 and 3 was to extend measurements to tripartite and quadripartite backgrounds, not to assess the effects of contrast per se, we report measurements only for the two bipartite displays that were directly comparable with the later tripartite and quadripartite cases.

### Method

The standard surround was divided vertically in half. One half was assigned a luminance of 30  $\text{cd}/\text{m}^2$ , and the other was assigned a luminance of 70  $\text{cd}/\text{m}^2$ . The position of each luminance was counter-balanced. The standard central surface could be either 25 or 75  $\text{cd}/\text{m}^2$ . The resulting two standards are reproduced in the insets of Figure 2. Each was paired with 9 different comparisons ranging from 10 to 90  $\text{cd}/\text{m}^2$  in steps of 10  $\text{cd}/\text{m}^2$ . Each observer made measurements on the 18 comparisons in randomized order.

### Results

Figure 2 plots the proportion of observers that judged each standard center (see insets) to be the same as or lighter than the corresponding comparisons as a function of the luminance of the comparison annuli. Here, and in the next two experiments, these proportions are analyzed to estimate values that function as an equivalent background, if any. Equivalent backgrounds are determined as the values of annulus luminance that yield a point of subjective equality (PSE) for the two center surfaces (50% *lighter* judgments corresponding to peak *same* judgments). No statistics are computed at this stage. In the Dis-

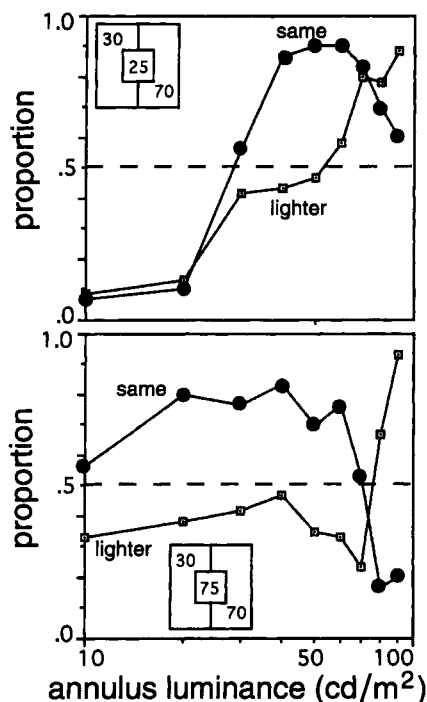


Figure 2. Proportions of *same* (circles) and *lighter* (squares) judgments for 25- and 75- $\text{cd}/\text{m}^2$  standard centers in front of bipartite surrounds, compared haploscopically with equal-luminance centers presented in the center of a single annulus. The abscissa represents the luminance of the single annulus.

cussion section, overall results from all three experiments are evaluated statistically against theoretical predictions.

Consider first the top graph of Figure 2 and the *lighter* judgments. When the comparison annulus luminance was low, essentially no observer judged the 25- $\text{cd}/\text{m}^2$  standard center to be lighter than the comparison. This makes sense, since this standard center is a decrement relative to both its surrounding luminances, whereas the 25- $\text{cd}/\text{m}^2$  comparison is an increment relative to low values of the annulus luminance (e.g., 25/10 for the first data point). Conversely, when annulus luminance was 70  $\text{cd}/\text{m}^2$  or higher, essentially all observers judged the standard to be lighter. This is also understandable, because the comparison center is now a larger decrement relative to annulus luminance (e.g., 25/90 for the last data point) than the standard center relative to either one of its surround patches (e.g., either 25/30 or 25/70 for the last data point). Finally, when the annulus luminance was between 30 and 60  $\text{cd}/\text{m}^2$ , the observers were split. The PSE for the center surfaces is found at the value yielding maximum uncertainty (50% *lighter* judgments), which is somewhere between 30 and 60  $\text{cd}/\text{m}^2$ . Next, consider the *same* judgments in the top graph of Figure 2. When annulus luminance was low, essentially no observer judged the standard center to be the same as the comparison. As annulus luminance approached 40  $\text{cd}/\text{m}^2$ , however, the proportion of *same* judgments approached 100%, reach-

ing a peak around 50–60 cd/m<sup>2</sup>. The position of this peak again suggests that the 25-cd/m<sup>2</sup> standard appeared equal to the 25-cd/m<sup>2</sup> comparison when it was surrounded by an annulus between 50 and 60 cd/m<sup>2</sup>. By linear interpolation, the PSE for the center surfaces is found to be at a 53-cd/m<sup>2</sup> comparison annulus. We may conclude that, for this type of display, there is an equivalent background in the sense that the effect of the 53-cd/m<sup>2</sup> annulus on the lightness of the comparison appears to be equivalent to the effect of the 30 and 70 patches on the standard.

Comparing these results with those in the bottom graph of Figure 2, consider the *lighter* judgments first. At low to intermediate annulus luminances, there is a weak bias against judging the standard lighter than the comparison, but the proportions exhibit no clear majority. As soon as the annulus luminance is increased from 70 to 80 cd/m<sup>2</sup>, however, essentially all observers judge the standard to be lighter. Applying the same logic as before, we may conclude that the PSE for the two center surfaces is located somewhere between 70 and 80 cd/m<sup>2</sup>. By linear interpolation, this point is 76 cd/m<sup>2</sup>.

The interpretation of this number, however, is different from that of the top graph of Figure 2. Consider the distribution of the *same* judgments. Over a wide range of annulus luminances (20 to 60 cd/m<sup>2</sup>), about 80% of the observers judged the two center surfaces to be the same lightness. Only after the annulus luminance increased from 70 to 80 cd/m<sup>2</sup> did the majority of the observers judge the centers to be different. This pattern of results may be interpreted as follows. As long as the two center surfaces are both increments (relative to the annulus luminance for the comparison, or relative to the two patches for the standard), there is always a very strong tendency for them to have the same lightness. This tendency is only slightly counteracted by a weak bias to see the comparison as lighter, perhaps due to the fact that the ratio of the comparison center to its annulus luminance is always larger than at least one of the ratios of the standard center to its surroundings and, in many cases, larger than both (e.g., 75/20 vs. 75/30 and 75/70). As soon as the comparison center becomes a decrement, all observers perceive it as darker than the standard. Thus, contrary to the top graph of Figure 2, there is no single annulus that works as an equivalent background in this display. Rather, there is a wide range of annuli, roughly corresponding to those that render the comparison center an increment, that yield judgments of equality between the two centers.

These results are in good qualitative agreement with those reported by Schirillo and Shevell (1993). For a surface that is a decrement relative to both surrounding patches, there is one value that functions as an equivalent background (about 53 cd/m<sup>2</sup>), and this is not too far from the space-average expectations. In fact, the space (geometric) average of the bipartite surround is about 46 cd/m<sup>2</sup>, only about 1/20 of a log unit from the observed PSE. The highest luminance prediction (70 cd/m<sup>2</sup>), conversely, is about 1/10 of a log unit from the observed PSE. (It may be noted, however, that if the space average is computed to include the center surface, then the expected

value shifts down to about 41 cd/m<sup>2</sup>, which is about as far from the PSE as the highest luminance expectation, although in the opposite direction.) For a surface that is an increment relative to both surrounding patches, the pattern of data is far from the expectations based on the space-average luminance hypothesis. With this display, there is no single annulus that functions as an equivalent background. Rather, several annuli do, as long as their luminance is lower than the center luminance. This is consistent with the highest luminance hypothesis. Within each monocular image, the center surfaces are both the highest luminance. On the basis of the highest luminance hypothesis, the center surfaces should both become white and therefore appear to have the same lightness.

## EXPERIMENT 2

In Experiment 2, equivalent background values for tripartite surrounds were determined. These surrounds were comparable to those used by Bruno (1992) in that they consisted of three different luminances. The methodology used was the same as that used in Experiment 1, except for the number of patches in the standard surround and the number of luminances of the central squares. Three luminances (30, 50, and 70 cd/m<sup>2</sup>) were randomly assigned to three sectors surrounding the standard central square, which had one of four different luminances. The 25-cd/m<sup>2</sup> standard center was the same as that in Experiment 1 and a decrement relative to all three sectors. The 75-cd/m<sup>2</sup> standard center was also the same as that in Experiment 1 and an increment relative to all three sectors. A 40-cd/m<sup>2</sup> standard center was a decrement to two surround surfaces and an increment to the third; the reverse was true for the 60-cd/m<sup>2</sup> standard center.

### Method

The four different standard configurations employed in this experiment are depicted in the insets of Figure 3. The factorial combination of the 4 central squares (25, 40, 60, or 75 cd/m<sup>2</sup>) and 9 comparison surrounds (10, 20, 30, 40, 50, 60, 70, 80, or 90 cd/m<sup>2</sup>) yielded a total of 36 displays. Each observer saw each of the 36 displays once in random order. Before each trial, the three luminances (30, 50, and 70 cd/m<sup>2</sup>) were randomly assigned to the three sectors surrounding the standard target.

### Results

Figure 3 plots the proportion of observers that judged each standard center (see insets) to be the same as or lighter than the corresponding comparisons as a function of the luminance of the comparison annuli. As in Experiment 1, equivalent backgrounds for each graph may be derived by computing values that yield exactly 50% *lighter* judgments. For the convenience of the reader, these are all summarized in Figure 4. However, as in Experiment 1, comparing the psychometric curves of the *lighter* judgments and the distributions of the *same* judgments suggests that these numbers do not have all the same interpretation.

Consider first the top left graph of Figure 3. The shape of the two curves is similar to that of the top graph of Fig-

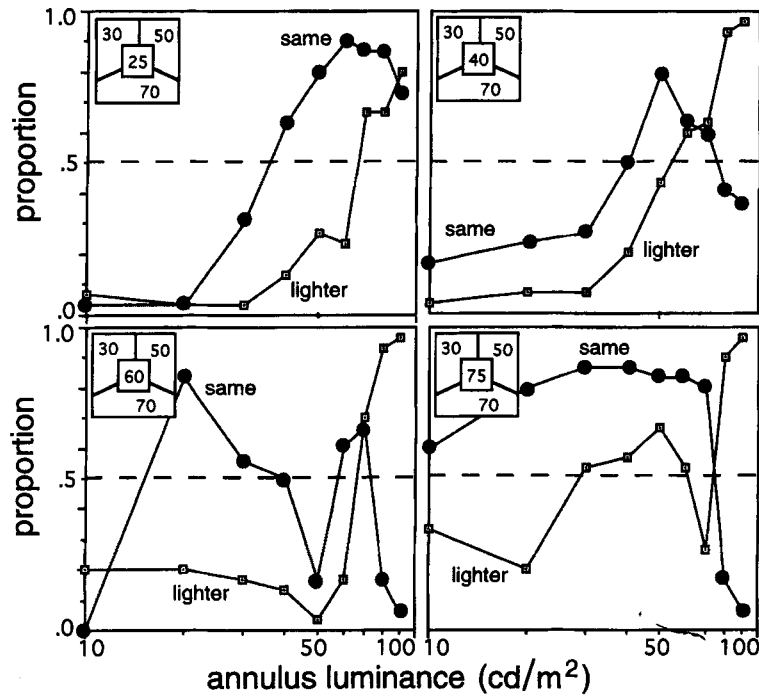


Figure 3. Proportions of *same* (circles) and *lighter* (squares) judgments for 25-, 40-, 60-, and 75-cd/m<sup>2</sup> standard center surfaces in front of tripartite surrounds, compared haploscopically with equal-luminance centers presented in front of a single annulus. The abscissa represents the luminance of the single annulus.

ure 2 (for Experiment 1). The psychometric curve of the *lighter* judgments crosses 50% between 60 and 70 cd/m<sup>2</sup> (the estimated value is 66 cd/m<sup>2</sup>). The distribution of the *same* judgments has a peak at 60 cd/m<sup>2</sup>. Therefore, there is one single surround luminance that equates the lightnesses of the two center surfaces. In the bottom right graph of Figure 3, the shape of the two curves is similar to the bottom graph of Figure 2 (for Experiment 1). The psychometric curve of the *lighter* judgments crosses 50% when the comparison surface becomes a decrement relative to the annulus, between 70 and 80 cd/m<sup>2</sup>. How-

ever, as in Experiment 1, there is a wide range of annulus luminances that yield *same* judgments. These include essentially all annuli lower than 75 cd/m<sup>2</sup>. Consistent with these judgments, *lighter* judgments also cross 50% in other positions (e.g., between 20 and 30 cd/m<sup>2</sup>). As in Experiment 1, therefore, there is no single equivalent background value in this type of display. Rather, it seems that several annuli can function as an equivalent background, as long as they are lower than 75 cd/m<sup>2</sup>.

Lastly, consider the top right and the bottom left graphs of Figure 3. These displays were fundamentally different from those in Figure 2 (for Experiment 1), in that their center surface had an intermediate luminance relative to the three surrounding patches. Nonetheless, the overall shape of the curves appears to be influenced by the qualitative relationships that prevail in the configuration. For instance, in the top right graph of Figure 3, the center surface is a decrement relative to two of the three surrounding patches. Therefore, this type of configuration is similar to that on the top left of Figure 3, and the pattern of data reflects this similarity. The psychometric curve of the *lighter* judgments crosses 50% only once (at 52 cd/m<sup>2</sup>), and this point closely corresponds with the peak of the *same* judgments (50 cd/m<sup>2</sup>). Therefore, for the top right configuration of Figure 3, there seems to be one single value that equates the lightnesses of the two center surfaces in this configuration. The main difference, relative to the top left configuration of Figure 3, is that this value is now lower by about 15 cd/m<sup>2</sup>.

		center luminance			
		25	40	60	75
standard surround	□□	53			76
	⊠	66	52	66	74
	⊞	69	48	67	71

Figure 4. Estimated equivalent background values for all displays of the three experiments. Values are luminances in candelas per square meter (cd/m<sup>2</sup>).

In the bottom left graph of Figure 3, the center surface is an increment relative to two of the three surrounding patches. Therefore, this type of configuration is most similar to the bottom right graph of Figure 3. This similarity is also reflected in the pattern of the data, although not as clearly as in the previous case. The psychometric curve of the *lighter* judgments also crossed 50% only once (at 66 cd/m<sup>2</sup>). However, this point does not correspond with the peak of the *same* judgments (which is located at 20 cd/m<sup>2</sup>). Additionally, *same* judgments in the range 20–60 cd/m<sup>2</sup> are all above 50%, with the exception of the trough at 50 cd/m<sup>2</sup>. Thus, as for the bottom right graph of Figure 3, the observed pattern is evidence that there is no single equivalent background. The most puzzling feature of the data represented in the bottom left graph of Figure 3 is the lack of consistency between the *same* judgments, peaking at a relatively low comparison surround, and the *lighter* judgments, which cross 50%—way above the space-average expectations. In the Discussion section, we provide a hypothesis that accounts for the pattern exhibited by these displays as well as those that share with them the feature of having a standard center surface with intermediate luminance relative to the standard surround luminances.

As reported by Bruno (1992), the present results tend to be more consistent with the highest luminance rule than with averaging. When the center surface is a decrement relative to all three surround patches, there is a single equivalent background value, and this is rather close to the highest luminance of the three. When the center surface is an increment relative to all three patches, several annuli function as an equivalent background, as one would expect based on a highest luminance rule. When the center luminance is intermediate relative to the three patches, the percept seems to be a compromise between two opposing tendencies—one for increments that prevails in the two-increment configuration, and one for decrements that prevails in the two-decrement case. As anticipated in the previous paragraph, further analysis of the possible nature of this compromise is deferred to the Discussion section.

### EXPERIMENT 3

In Experiment 3, equivalent background values for quadripartite surrounds were determined. The methodology was the same as that of Experiments 1 and 2, except for the number of patches in the standard surround and the number of central surfaces. Four luminances (30, 45, 55, and 70 cd/m<sup>2</sup>) were randomly assigned to the four square patches surrounding a standard center square, which could have one of four different luminances. These were the same as in Experiment 2. Thus, the 25-cd/m<sup>2</sup> standard center was a decrement relative to all four surrounding patches, the 75-cd/m<sup>2</sup> standard center was an increment relative to all four patches, the 40-cd/m<sup>2</sup> standard center was a decrement to three surfaces and an increment to the fourth, and the 60-cd/m<sup>2</sup> standard center

was an increment to three surfaces and a decrement to the fourth.

### Method

The four standard configurations employed in Experiment 3 are depicted in the insets of Figure 5. The factorial combination of the 4 central squares (25, 40, 60, and 75 cd/m<sup>2</sup>) and 9 comparison surrounds (10, 20, 30, 40, 50, 60, 70, 80, and 90 cd/m<sup>2</sup>) yielded a total of 36 displays. Each observer saw each of the 36 trials once in random order.

### Results

Figure 5 plots the proportion of observers that judged each standard center (see insets) to be the same as or lighter than the corresponding comparisons as a function of the luminance of the comparison annuli. As in Figures 2 and 3, equivalent backgrounds for each graph may be derived by computing values that yield exactly 50% *lighter* judgments (see Figure 4) and then comparing these with the peaks of the *same* judgment distributions.

The results are similar to those of Experiment 2. When the standard center is a decrement relative to all four surrounding patches, a single annulus value functions as an equivalent background. The psychometric curve of the *lighter* judgments crosses 50% only once (at 69 cd/m<sup>2</sup>), and this value closely corresponds to the peak of the *same* judgments (70 cd/m<sup>2</sup>). When the standard center is an increment relative to all four surrounding patches, several annuli function as equivalent backgrounds. The psychometric curve of the *lighter* judgments crosses 50% three times, and all annuli below 75 cd/m<sup>2</sup> yield between 80% and 90% *same* judgments. Finally, the two standards having intermediate luminances relative to their surrounding patches provide evidence for a compromise between matching strategies for increments and decrements. In the configuration with the 40-cd/m<sup>2</sup> standard center, decrements prevail and a single annulus appears to function as an equivalent background (at 48 cd/m<sup>2</sup> based on the *lighter* judgments, corresponding to a peak of *same* judgments at 40 cd/m<sup>2</sup>). In the configuration with the 60-cd/m<sup>2</sup> standard, increments prevail and there is a tendency to prefer a *same* judgment with all annuli below 60 cd/m<sup>2</sup>.

Again, these results are more in agreement with the highest luminance hypothesis than with schemes based on averaging. Three equivalent background values out of four are equal or almost equal to the highest luminance of the standard surround.

### DISCUSSION

Figure 6 summarizes the results of all three experiments. To compare the results against predictions based on the three candidate references, the data are presented in terms of contrast between the center surfaces and the corresponding observed equivalent backgrounds. Equivalent backgrounds were computed based on annulus values yielding 50% *lighter* judgments.

Predictions were derived as follows. For the space-average luminance hypothesis, in each configuration, we

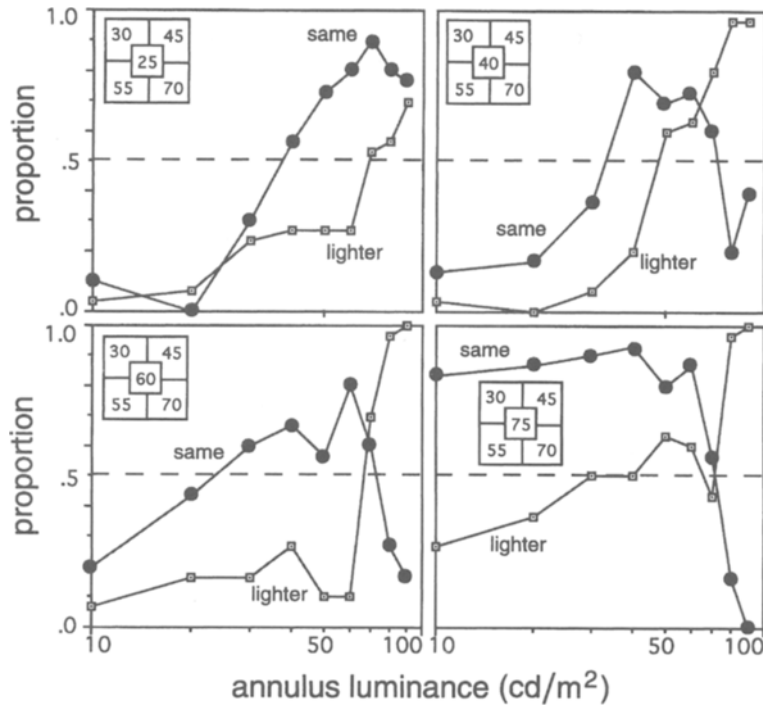


Figure 5. Same as Figure 3, for standard center surfaces presented in front of quadripartite surrounds.

computed the geometric average of all luminances, including the center. The Michelson contrast of the center surface luminance to this geometric average is the prediction based on space-average luminance. For the space-average contrast hypothesis, we computed Michelson contrasts of all surrounding patches with their center surfaces. The average of these is the prediction based on space-average contrast. Finally, for the highest luminance hypothesis, we simply computed the Michelson contrast of the center surface luminance to the highest luminance

in the configuration. When the center surface is itself the highest luminance, the prediction is zero contrast. Recall that the highest luminance hypothesis predicts that any annulus value lower than the comparison center should function as an equivalent background. On the basis of this hypothesis, therefore, *lighter* judgments should stay at 50%, as long as the annulus luminance functions as an equivalent background, and then cross steeply to 100%, as soon as the annulus luminance surpasses the center luminance.

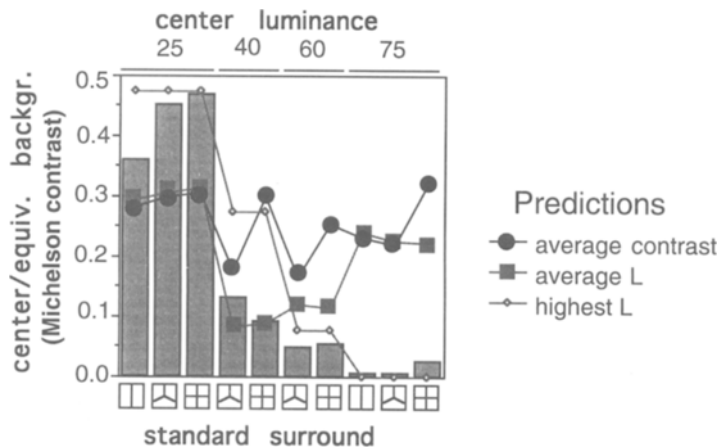


Figure 6. Observed ratios of center square luminances to their equivalent surrounds (gray bars), plotted with predictions based on three alternative anchoring schemes: space-average contrast (circles), space-average luminance (squares), and highest luminance (diamonds).



The highest luminance model provides the closest fit to the data for most of the observations. Tests of goodness of fit confirmed that the observed ratios were not statistically distinguishable from the highest luminance expectations [ $\chi^2(9) = 1$ ,  $p = .99$ ], whereas they were much less consistent from the other two sets of expected values [ $\chi^2_s(9) = 19$  and  $16$ ,  $ps < .03$  and  $.06$ , for the space-average luminance and the space-average contrast hypotheses, respectively]. Thus, the present results suggest that, in most conditions, a homogeneous surround is equivalent to a complex surround when the homogeneous surround is closest to the highest luminance. This finding is consistent with proposals that the human visual system uses the highest luminance of a scene as the reference value for mapping local ratios to lightness values (Cataliotti & Gilchrist, 1995; Land & McCann, 1971; Wallach, 1948) and does not provide evidence in favor of computational models that use spatial averaging. However, the highest luminance predictions do not fit the data perfectly. Two features are of chief interest here: the change in the observed equivalent backgrounds as the complexity of the standard surround increased from two to four patches, and the values observed in the configurations having an intermediate center luminance relative to the surrounding patches.

### Surround Complexity

The highest luminance hypothesis, as well as each of the other hypotheses that we now reject, is a photometric anchoring process. The reference term is computed from the highest luminance in the scene, independent of its spatial distribution. However, there are several reasons to believe that the anchoring process is not blind to the spatial properties of the image. For instance, spatial factors can control the direction of contrast in some classic illusions, such as the Benary effect (Benary, 1924). The importance of spatial complexity in providing a well-defined perceived gray scale is well known (Arend & Spehar, 1993a, 1993b). In the displays investigated here, the spatial complexity of the standard surround, defined simply as the number of patches, also appears to have an effect on the observed equivalent backgrounds.

Consider the equivalent background values for the 25-cd/m<sup>2</sup> central squares in Experiment 1. On the basis of the highest luminance rule, the bipartite surrounds of Experiment 1 should behave like the tripartite and quadripartite stimuli in Experiments 2 and 3. The highest luminance is 70 cd/m<sup>2</sup> in all three configurations. Therefore, the reference used by the integration process should also be the same in all three configurations. However, in Experiment 1, equivalent background values were substantially lower than the highest luminance rule would predict. In fact, the tripartite surrounds were also somewhat lower than the highest luminance. It was only with the quadripartite surround that equivalent backgrounds were approximately similar to the highest luminance rule. This trend is consistent with an effect of spatial complexity.

Next, consider when the center surfaces were 75 cd/m<sup>2</sup>. In all these displays, the standard centers were the highest luminance in each monocular image. Thus, given the highest luminance rule, they should have all appeared white. However, all the comparison centers surrounded by annuli with luminances lower than 75 cd/m<sup>2</sup> were also the highest luminance of their configurations. Thus, all of the incremental comparison centers should also have appeared white. In other words, the highest luminance rule predicts that all 75-cd/m<sup>2</sup> comparison centers should have appeared equally light to the 75-cd/m<sup>2</sup> standard center surface, when the comparison centers were also increments relative to their surrounds. Thus, the proportion of *same* judgments should be a step function of surround luminance, yielding 100% *same* judgments when the comparison center is an increment and dropping to 0% when it becomes a decrement. Likewise, the rule would predict that the *lighter* judgments remain at 50% when the comparison center is an increment and jump to 100% when it becomes a decrement. Although the data were basically consistent with these predictions, preferences for a *same* forced choice fell short of 100%. Concurrently, the proportions of *lighter* judgments fell below 50%. This deviation from the expected pattern based on the highest luminance hypothesis was largest in Experiment 1, somewhat smaller in Experiment 2, and smallest in Experiment 3. Again, it seems that the data vary as a function of spatial complexity.

Thus, it seems that, as the spatial complexity of the standard configuration increases, the pattern of results becomes more similar to the expectation based on the highest luminance hypothesis. This effect may account for the discrepancy between the results of Bruno (1992), which supported the highest luminance hypothesis, and those of Schirillo and Shevell (1993), which were closer to the average luminance hypothesis for decrements. Bruno used Mondrian surrounds consisting of three different luminances distributed over six different patches. Schirillo and Shevell used checkerboards consisting of two different luminances distributed over four patches. The two configurations are fairly similar in their spatial complexity. Yet, on the basis of the results of the experiments reported here, increasing the number of surround luminances from two to three seems sufficient to move observed equivalent backgrounds much closer to the highest luminance expectations.

We know of only two approaches to the anchoring problem that attempt to account for spatial factors. The first is Li and Gilchrist's (1993) proposal of an *area-luminance* hypothesis, which pits two opposing tendencies against each other to determine which value is chosen as the effective reference for a given scene. One tendency is to use the highest luminance, whereas the second is to use the largest area. When the highest luminance region is also the larger area, the choice is straightforward. Otherwise, the two tendencies conflict. If the darker region is much larger than the highest luminance region, then the

larger region may appear white, while the highest luminance region appears luminous. If the difference in size is less dramatic, the highest luminance may appear white and the darker region takes on a lightness proportional to its relative area.

Strictly speaking, Li and Gilchrist's (1993) proposal applies only to two-luminance displays filling the entire field of view. At the distance employed in our study, the monitor used to display our configurations filled about  $40^\circ \times 30^\circ$  of visual angle. Under these conditions, effects from the remaining, uncontrolled, part of the field of view have been shown to be measurable, but minimal (Agostini & Bruno, 1996). However, the area-luminance rule of Li and Gilchrist does not seem to predict the pattern of our data. For instance, the area of highest luminance (in both cases,  $70 \text{ cd/m}^2$ ) was larger in the bipartite surround than in the quadripartite one. Therefore, on the basis of the area-luminance hypothesis, the highest luminance patch should have had greater weight than the enclosing background in the bipartite case, relative to that in the quadripartite one. However, the opposite trend is observed in the data: Equivalent background values were essentially identical to the highest luminance in the quadripartite configuration and were substantially lower than the highest luminance in the bipartite configuration.

The second approach to anchoring is adopted by the early version of the *Retinex* model (Land, 1977; Land & McCann, 1971). In this version (sometimes called *Retinex-with-reset*), lightness values are computed for each patch by several parallel computations, which sequentially multiply luminance ratios over space. Whenever the product of the multiplication becomes larger than 1, the process has encountered a surface of higher luminance than the previous one, and the computation is restarted. The effect of this reset operation is to assign a lightness value to each surface based on the ratio with the highest encountered luminance. Lightnesses computed by parallel spatial processors are then averaged to obtain a final value. Although the convergence properties of the *Retinex-with-reset* algorithm are not well understood (Brainard & Wandell, 1986), it seems clear that final lightness values depend on two parameters: the number of parallel spatial processes, and their spatial extent. Assuming that these would be influenced by spatial properties of the image, such as the number of edges in the image, it is plausible that a modification of the *Retinex-with-reset* algorithm would predict the present pattern of data.

Given the observed effect of complexity, further research on equivalent backgrounds should explore potential effects of using standard surrounds containing more patches at different spatial scales, as well as a larger range of luminances. In a natural image, the typical range of luminances for surfaces that appear to span the whole lightness range from black to white is about 1:30, whereas, in the displays investigated here, this range was restricted to only 1:3 ( $25\text{--}70 \text{ cd/m}^2$ ) or even smaller in some displays (e.g.,  $30\text{--}75 \text{ cd/m}^2$ ). It is presently unknown whether increasing the range of luminances in the standard sur-

round would affect equivalent background computations. In addition, natural images tend to contain luminance information at several spatial scales that is not conveyed in simplified surrounds such as those investigated here. These differences may account for the results reported by Valberg and Lange-Malecki (1990), whose reported equivalent backgrounds were consistent with spatial averaging. However, they used a chromatic Mondrian containing several patches of different colors. Their surround was therefore significantly more complex than those of Bruno or Schirillo and Shevell. It is possible that integration mechanisms operating at different spatial scales come into play under these conditions. Additionally, their use of chromatic patches renders it difficult to make a direct comparison with the other two studies. However, the issues of equivalent backgrounds for displays including patches at different spatial scales and for chromatic surrounds are certainly important and deserve further study.

### Intermediate Center Surfaces

Substantial deviations from the highest luminance expectations were found with the  $40\text{-cd/m}^2$  center squares. We propose that the behavior of this configuration may result from a compromise between different matching strategies for increments and decrements. For example, consider the shape of the *same* distribution for the  $40\text{-cd/m}^2$  center square surrounded by four patches. This center square is an increment relative to the  $30\text{-cd/m}^2$  patch. Therefore, relative to this patch, the standard center should appear to have the same lightness as the comparison center as long as it is also an increment relative to the comparison annulus. As soon as the comparison surround becomes greater than  $40 \text{ cd/m}^2$ , however, the standard center should appear lighter than the comparison center. This theoretical function would have the same shape as the step function of the  $75\text{-cd/m}^2$  center squares, although it would step down at  $40 \text{ cd/m}^2$  rather than  $75 \text{ cd/m}^2$ . On the other hand, the  $40\text{-cd/m}^2$  center square is a decrement relative to all the other patches. Therefore, relative to all these patches, the standard center should appear to have the same lightness as the comparison center only when this is surrounded by the annulus that corresponds to the highest luminance of the four patches. This theoretical function would have the same shape as that seen for the  $25\text{-cd/m}^2$  center squares. Because the two theoretical functions overlap, it is possible to weight and combine them so as to produce a distribution of *same* responses that peaks approximately at  $50 \text{ cd/m}^2$  and approximately resembles the observed one (see Figure 7). Similar logic can be applied to the  $60\text{-cd/m}^2$  center squares, which are also partly incremental and partly decremental relative to their surrounding patches. In this case, however, one should assume that the theoretical increment function steps down at  $60 \text{ cd/m}^2$ . This may account, after the appropriate weights are chosen, for the fact that the observed curves are more similar to the highest luminance predictions in this case.

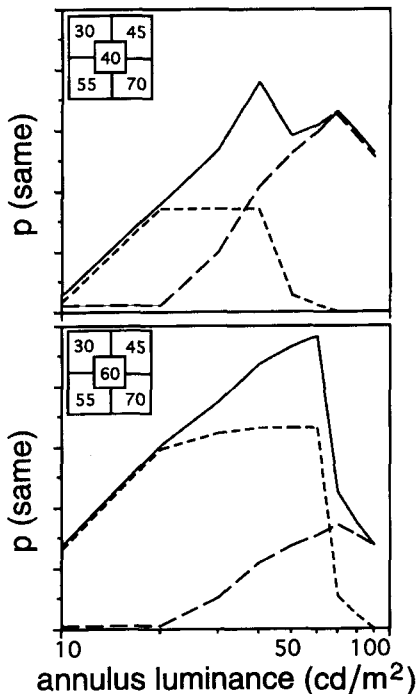


Figure 7. Illustration of how the shape of the *same* distributions for Experiment 3 (continuous line) could be approximately obtained by combining functions for increments (shorter dashes) and decrements (longer dashes) and weighting them appropriately. The two functions are scaled to reflect the weights used.

## Conclusion

In conclusion, consider again our initial question. A surface made of gray paper is surrounded by many other gray papers. Another surface, cut from the same gray paper, is surrounded by a single gray paper. If you want the two papers to be seen as identical, what shade should you choose for the single gray surround? Overall, it seems that the best shade to choose is the lightest of the several surfaces surrounding the other center surface. In general, this will produce a better match than predictions based on spatial averaging of luminance or contrast. However, it will not always produce a perfect match. Two main factors appear to provoke these deviations from the highest luminance predictions: the spatial complexity of the display, and the qualitative relationship (incremental or decremental) of a given surface to its immediate surrounds.

## REFERENCES

- AGOSTINI, T., & BRUNO, N. (1996). Lightness contrast in CRT and paper-and-illuminant displays. *Perception & Psychophysics*, **58**, 250-258.
- AREND, L. E. (1985). Spatial gradient illusions and inconsistent integrals. *Investigative Ophthalmology & Visual Science*, **26**(Suppl.), 280.
- AREND, L. E., & GOLDSTEIN, R. (1987). Simultaneous constancy, lightness and brightness. *Journal of the Optical Society of America A*, **4**, 2281-2285.
- AREND, L. E., & SPEHAR, B. (1993a). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics*, **54**, 446-456.
- AREND, L. E., & SPEHAR, B. (1993b). Lightness, brightness, and bright-

- ness contrast: 2. Reflectance variation. *Perception & Psychophysics*, **54**, 457-468.
- BENARY, W. (1924). Beobachtungen zu einem Experiment über Helligkeitskontrast. *Psychologische Forschung*, **5**, 131-142. [Also published, 1938, as The influence of form on brightness contrast. In W. D. Ellis (Ed.), *A source book for Gestalt psychology* (pp. 104-108). London: Routledge & Kegan Paul]
- BLAKE, A. (1985). Boundary conditions for lightness computations in Mondrian world. *Computer Vision, Graphics & Image Processing*, **32**, 314-327.
- BRAINARD, D. H., & WANDELL, B. A. (1986). Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, **3**, 1651-1661.
- BROWN, R. (1994). The world is not gray. *Investigative Ophthalmology & Visual Science*, **35**(Suppl.), 2165.
- BROWN, R., & MACLEOD, D. (1991). Induction and constancy for color saturation and achromatic contrast variance. *Investigative Ophthalmology & Visual Science*, **32**(Suppl.), 2673.
- BRUNO, N. (1992). Lightness, equivalent backgrounds, and the spatial integration of luminance. *Perception*, **21**(Suppl.), 80.
- CATALIOTTI, J., & GILCHRIST, A. [L.] (1995). Local and global processes in surface lightness perception. *Perception & Psychophysics*, **57**, 125-135.
- CRAWFORD, B. H. (1947). Visual adaptation in relation to brief conditioning stimuli. *Proceedings of the Royal Society of London: Series B*, **134**, 283-302.
- GEISLER, W. S., & ALBRECHT, D. G. (1992). Cortical neurons: Isolation of contrast gain control. *Vision Research*, **32**, 1409-1410.
- GILCHRIST, A. L. (1988). Lightness contrast and failures of constancy: A common explanation. *Perception & Psychophysics*, **43**, 415-424.
- GILCHRIST, A. [L.], DELMAN, S., & JACOBSEN, A. (1983). The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception & Psychophysics*, **33**, 425-436.
- HELSON, H. (1943). Some factors and implications of color constancy. *Journal of the Optical Society of America*, **33**, 555-567.
- HELSON, H. (1964). *Adaptation level theory*. Harper & Row.
- HORN, B. K. P. (1974). Determining lightness from an image. *Computer Graphics & Image Processing*, **3**, 277-299.
- HURLBERT, A. [C.] (1986). Formal connections between lightness algorithms. *Journal of the Optical Society of America A*, **3**, 1684-1693.
- HURLBERT, A. C., & POGGIO, T. A. (1988). Synthesizing a color algorithm from examples. *Science*, **239**, 482-485.
- JUDD, D. B. (1940). Hue saturation and lightness of surface colors with chromatic illumination. *Journal of the Optical Society of America*, **30**, 2-32.
- KATZ, D. (1911). Die Erscheinungsweisen der Farben und ihre Beeinflußung durch die individuelle Erfahrung [The appearance of colors and its dependence on the individual experience]. *Zeitschrift für Psychologie*, **7**(Suppl.).
- KNAU, H., & SPILLMANN, L. (1994). Nulling of brightness fading in a Ganzfeld. *Perception*, **23**(Suppl.), 107.
- LAND, E. (1977). The retinex theory of color vision. *Scientific American*, **237**, 108-128.
- LAND, E. (1986). Recent advances in retinex theory. *Vision Research*, **26**, 7-22.
- LAND, E., HUBEL, D. H., LIVINGSTONE, M. S., PERRY, S. H., & BURNS, M. M. (1983). Colour-generating interactions across the corpus callosum. *Nature*, **303**, 616-618.
- LAND, E., & MCCANN, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America*, **61**, 1-11.
- LI, X., & GILCHRIST, A. L. (1993). Geometric configuration and anchoring of surface lightness. *Investigative Ophthalmology & Visual Science*, **34**(Suppl.), 748.
- MCCANN, J. (1989). The role of simple nonlinear operations in modeling human lightness and color sensations. *SPIE Proceedings*, **1077**, 355-363.
- MCCANN, J. (1992). Rules for colour constancy. *Ophthalmological & Physiological Optics*, **12**, 175-177.
- ROBSON, J. G. (1991). Neural coding of contrast in the visual system [Abstract]. *Optical Society of America Technical Digest Series*, **17**, 152.

- SCHIRILLO, J. A., & SHEVELL, S. K. (1993). Brightness contrast from a complex background requires a complex description. *Investigative Ophthalmology & Visual Science*, **34**(Suppl.), 746.
- SCHUBERT, J., & GILCHRIST, A. L. (1992). Relative luminance is not derived from absolute luminance. *Investigative Ophthalmology & Visual Science*, **33**(Suppl.), 1258.
- VALBERG, A., & LANGE-MALECKI, B. (1990). "Colour constancy" in Mondrian patterns: A partial cancellation of physical chromaticity shifts by simultaneous contrast. *Vision Research*, **30**, 371-380.
- WALLACH, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, **38**, 310-324.
- WHITTLE, P., & CHALLANDS, P. D. C. (1969). The effect of background luminance on the brightness of flashes. *Vision Research*, **9**, 1095-1110.

**NOTE**

1. Probit analysis, the standard technique for obtaining a PSE in a psychometric function, could not be applied, since it assumes that distributions of differential judgments are cumulative normal sigmoids. As will be seen in the plots of the results, some of our data exhibited massive violations of this assumption in certain conditions. In any event, the reader can be assured that, for curves that were more similar to cumulative normal sigmoids, probit analysis yielded essentially the same PSE values as straightforward linear interpolation.

(Manuscript received March 18, 1996;  
revision accepted for publication July 18, 1996.)