# A Victory for Equivalent Background — On Average

*Mark D. Fairchild*
*Munsell Color Science Laboratory*
*Rochester Institute of Technology*
*Rochester, New York, USA*

## Abstract

A psychophysical experiment was carried out to examine the relationship between image contrast and overall perceived brightness. A second phase of the experiment looked at the relationship between the perceived brightness of variegated backgrounds and the simultaneous contrast effect produced by such backgrounds. These results have important ramifications for procedures used to calculate adapting chromaticities and luminances for image displays. The results suggest that the traditional concepts of linear luminance integration and equivalent background are satisfactory on average. However, results for individual observers show very striking, consistent, and significant trends with substantial inter-observer variability. These results help to reconcile differences between fundamental vision science experiments and practical experiences with color appearance models.

## Introduction

It is well established that as the overall luminance level of a scene increases the scene appears to increase in contrast. This phenomenon has been referred to as the Stevens effect and has been incorporated in a variety of color appearance models and image reproduction processes.[1] An interesting and related phenomenon has also been reported informally. This is the observation that as the contrast of a scene or image increases at constant luminance, the apparent brightness of the scene will increase. This sort of phenomenon is often used to explain why a scene appears brighter through ski goggles, or sunglasses, with yellow lenses even though the luminance reaching the eye has decreased (scene contrast increases). This phenomenon was reported approximately 50 years ago by the OSA as quoted below.

> *The potency of this influence of comparison in perception is well illustrated by the illusion of heightened luminance in scenes where brightness differences are large, and the illusion of lowered luminance in scenes where the brightness differences are small. As a consequence of this effect, which leads to erroneous judgements of scene luminance, photographers sometimes unintentionally underexpose a "contrasty" theatrical scene indoors but overexpose a dull flat scene outdoors. (p. 154)*[2]

The experiments described in this paper were designed to quantitatively examine the relationship between brightness and image contrast at constant luminance and the impact of these effects on the appearance of image elements.

This work was directly motivated by the results of Oskoui and Pirrotta presented at the sixth Color Imaging Conference.[3] They showed that the adapted white point on CRT displays varied as a function of the contrast distribution of the adapting background despite constant average luminance and chromaticity. The current research was undertaken with the aim of better understanding how observers integrate a variegated background (*i.e.*, image) to establish an average perceived brightness and color. The hope was that a nonlinear (presumably expansive) integration function could be established that would reconcile the Oskoui and Pirrotta results with the concept of equivalent backgrounds used to establish adaptation points in typical color appearance models.

Others have reported similar results, but the various interpretations are not consistent. Brown and MacLeod[4] showed that various colored stimuli that appeared quite chromatic on a uniform gray background would all appear nearly achromatic on a variegated background with high luminance and chromatic contrast. They concluded that their results indicated some form of simultaneous contrast in the contrast, rather than luminance, domain. This is similar to a form of contrast adaptation or contrast gain control. In a series of papers,[5-8] Zaidi and coworkers examined brightness induction from uniform and complex surrounds and developed a model of contrast gain control to explain their results. Their work also indicated, for a small number of observers, that a variegated background of high contrast would induce gray patches to look lower in contrast (*i.e.*, dark patches look lighter than on a uniform background and light patches look darker than on a uniform background). Adelson[9] has illustrated quite different results, albeit with a different background configuration. Adelson showed that simultaneous contrast is enhanced, rather than diminished, on variegated backgrounds in comparison with uniform backgrounds of the same mean luminance. He interprets these results and various other observations using a so-called *apparent atmospheric transfer function* that is applied by the visual system at each point in an image to map luminance into perceived reflectance. The atmosphere can be characterized with a gain (change in level of illumination) and an offset (change in interposed transmittance, *e.g.* fog) which are compensated for in order to obtain perceived reflectance for various stimuli in a scene. This interesting and apparently robust interpretation can result in a form of contrast gain (perceived contrast increases with contrast) or contrast gain control (perceived contrast decreases with contrast). Schirillo and Shevell[10] found yet another type of results showing an increase in simultaneous contrast for stimuli with luminance above the integrated luminance of the background and no effect for stiumuli with luminance less than the integrated luminance of the background. They attempt to explain their results, for two observers, using various spatial vision models. It is important to note that the various interpretations are based on different types of stimulus configurations and observer tasks and thus might not be as contradictory as they seem upon first examination.

The hypothesis examined in the current research was that the perceived brightness of an image would increase with image contrast at constant luminance. If this is the case, then an expansive luminance integration function could be used to predict the perceived average brightness of various images. Further it was hoped that such a prediction could be used to predict the simultaneous contrast effects of various backgrounds and, by extension, the color appearance of image elements. The experiments detailed below were designed to evaluate these hypotheses.

## Experimental

The experimental images used as backgrounds consisted of 240x240-pixel regions made up of 12x12-arrays of 20x20-pixel squares. Six different contrast levels were used as illustrated in Fig. 1. Each square in a given background was assigned a gray level randomly (uniform distribution) from a set of four levels. The four levels associated with each image contrast are listed in table I. All luminance measurements are relative luminance where a 1.0 represents the maximum luminance of the display (97 cd/m$^2$ with an approximate D93 white point). Each background had an average relative luminance of 0.5 throughout the experiments. A Sony GDM-2000TC CRT display driven by a Power Macintosh G3/400 system was used throughout the experiments. The experimental stimuli and observer interface were generated and controlled using the IDL system. The display was characterized using standard colorimetric techniques and the accuracy of the characterization was evaluated by measuring the integrated luminance of a series of test backgrounds at all six contrast levels. The mean relative luminance of each of the backgrounds was 0.5 plus-or-minus 3%. The variance in the background relative luminance was due to the size of the integration aperture of the colorimeter and the random assignment of gray levels to each square for each measurement (analogous to a granularity measurement). The mean luminance for every background was never statistically significantly different from the aim of 0.5.

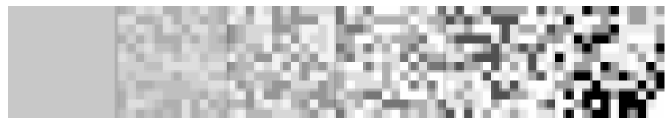

*Figure 1. Example backgrounds with contrasts 0f 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0 respectively. Each background integrates to a relative luminance of 0.5.*

**Table I. Relative luminance of each gray level for each image contrast.**

| Contrast | Level 1 | Level 2 | Level 3 | Level 4 |
|---|---|---|---|---|
| 0.0 | 0.500 | 0.500 | 0.500 | 0.500 |
| 0.2 | 0.400 | 0.467 | 0.533 | 0.600 |
| 0.4 | 0.300 | 0.433 | 0.567 | 0.700 |
| 0.6 | 0.200 | 0.400 | 0.600 | 0.800 |
| 0.8 | 0.100 | 0.367 | 0.633 | 0.900 |
| 1.0 | 0.000 | 0.333 | 0.667 | 1.000 |

Two types of experiments were run. In the first, observers were shown one of the background images and asked to use a slider to adjust a uniform area of the same size (approximately 4° angular subtense) to match in perceived brightness. Each contrast level was presented five times for a total of 30 trials. The trials were presented in random order and the spatial configuration of the background image was randomly generated for each trial. An example of the stimulus and interface configuration for this experiment is shown in Fig. 2. The starting luminance and slider-end-point values were also randomized for each trial such that observers could not learn an association between slider location and brightness over the course of the experiments.

The second type of experiment involved a simultaneous contrast measurement. A larger patch (approximately 1°, with relative luminance of either 0.4 or 0.6) was placed in the center of both the test and matching backgrounds. The observers' task then became to adjust the luminance of the uniform background such that the central patches matched in perceived lightness. Each patch (2) was presented on each background (6) five times for a total of 60 trials. Again the trials were presented in random order and the spatial configuration of the background image was randomly generated for each trial. An example of the stimulus and interface configuration for this experiment is shown in Fig. 3.

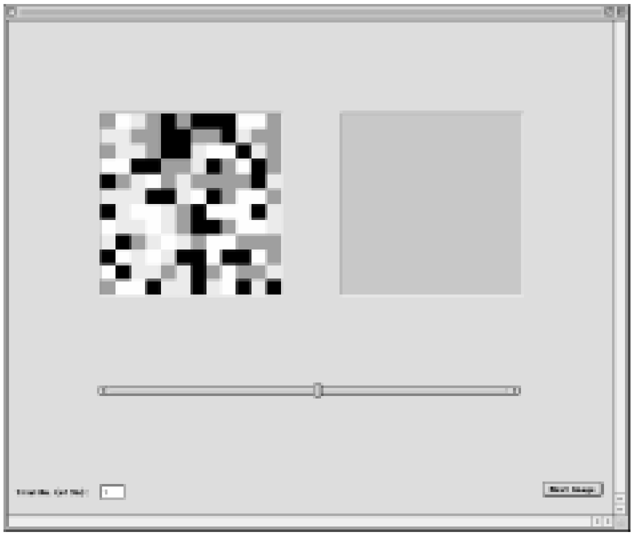

*Figure 2. Example configuration for the brightness matching experiment.*

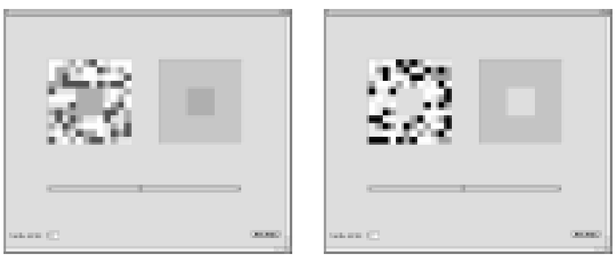

*Figure 3. Example configurations for the simultaneous contrast experiment. Left side for patch relative luminance = 0.4, right side for patch relative luminance = 0.6.*

Observers completed 6 practice trials prior to the 90 experimental trials and normally completed the full experimental task in approximately one hour. Seventeen observers, most experienced in color science and visual experiments, completed 18 sets of observations (one observer, the author, completed the experiment twice). The observers ranged in age from 23 to 40 years. The exact instructions given to the observers are presented below.

**INSTRUCTIONS**

You will be shown a stimulus configuration with two square fields. For each trial, the left field will be set to either a uniform gray or a pattern of gray squares. The right field will always be uniform. In some trials a larger gray square will be present in the middle of both fields.

**When no central gray squares are present (trials 7-36):**

You are to judge your impression of the average brightness of the left field and use the slider to adjust the brightness of the right field until it matches the average brightness of the left field. *(i.e., Match the brightnesses of the fields.)*

**When central gray squares are present (trials 37-96):**

You are to judge the brightness of the central gray square in the left field and use the slider adjust the background of the right field until the central gray square in the right field matches the brightness of the central gray square in the left field. *(i.e., Match the brightnesses of the central squares.)*

**Practice Trials:**

The first 6 trials are practice (3 without the squares followed by 3 with the squares). The experimenter will observe you during these trials to make sure you have properly understood the instructions. Please feel free to ask any questions during these first 6 trials.

**Remember:**

There is a total of 96 trials and the program will automatically exit upon completion. The slider location and the brightness levels for the slider end points are randomly reset at the beginning of each trial. Thus, the relationship between slider location and right-field brightness varies from trial to trial. When you've completed a match press the "*Next Image*" button to go on to the next trial.

The data collected were the relative luminance of the uniform background set by the observer for each trial. The overall results and examples for some individual observers are presented in the next section.

## Results and Discussion

Figures 4-6 show the results for all of the observers and each experimental phase. The average results are also shown as the thick black lines on each plot. Several points are evident from these results. First, when the test background is uniform (contrast = 0.0) observers make veridical luminance matches in both the brightness matching and simultaneous contrast tasks. As contrast increases, the matching luminance varies from a simple linear luminance integration for each observer. However, the trend seems to vary widely from observer to observer resulting in increasing variance in the results as background contrast increases.

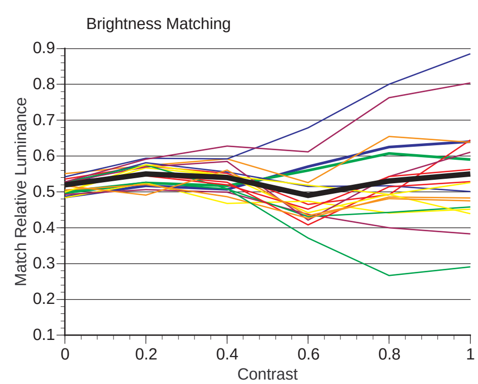


*Figure 4. Overall and mean results for the brightness matching task.*

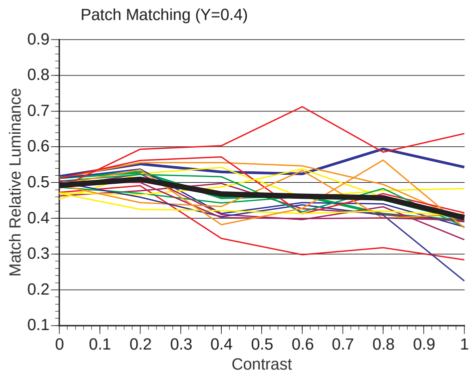


*Figure 5. Overall and mean results for the simultaneous contrast task with patch relative luminance of 0.4.*

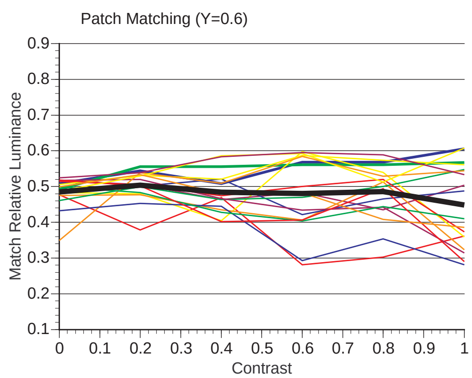


*Figure 6. Overall and mean results for the simultaneous contrast task with patch relative luminance of 0.6.*

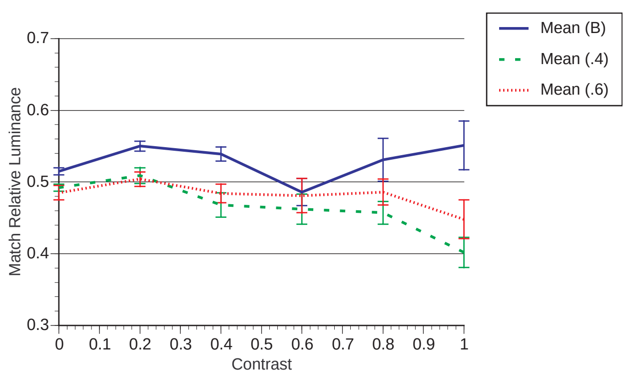


*Figure 7. Mean results for all three experimental phases. Error bars are plus-and-minus one standard error of the mean.*

The average results are examined in further detail in Fig. 7 showing the mean results across all 18 sets of observations for each of the three experimental phases together with error bars representing plus-and-minus one standard error of the mean. A general conclusion from Fig. 7 is that image contrast has little effect on perceived brightness. When the uncertainties are considered, only 4 data points on Fig. 7 are significantly different from 0.5 at a 95% confidence. These are the brightness matching results at contrasts of 0.2 and 0.4 and the simultaneous contrast results for the 0.4 relative luminance patch and background contrasts of 0.8 and 1.0. The brightness matching results show a trend toward an increase in brightness with contrast punctuated with a dip for the background with contrast of 0.6. This dip is present for a large number of observers and cannot be explained by any known experimental artifact. Perhaps it is due to the relationship between the test background and the window background of the experimental stimulus configuration. The window background had a relative luminance of just under (by one 8-bit digital count) 0.60, the relative luminance of level 3 in the 0.6 contrast backgrounds. At this contrast level, the background did seem to undergo some sort of change in viewing mode. The lower contrast backgrounds seemed to be behind some sort of fog or flare, while higher contrast backgrounds appeared to be illuminated by more light. This explanation is consistent with Adelson's apparent atmospheric transfer function and an increase in perceived brightness with contrast. The mean results for the induction experiments show a slight trend for the both patches to look lighter on the variegated backgrounds than on the uniform backgrounds (since a lower matching luminance was required for the uniform background controlled by the observers). This result is consistent with neither a contrast gain control nor a simple contrast gain. This result is not of too much concern since the trends are not really significant given the observer variability. The best conclusion to be drawn from Fig. 7 is that, on average, observers match a variegated background with a uniform background equal to the mean luminance (linear integration) and that simultaneous contrast can be predicted with an "equivalent background" model (similar to a "gray world" model). This conclusion contradicts the research cited in the introduction, but does explain the general success of traditional color appearance models that rely on the assumption that adaptation to a complex image is equivalent to adaptation to a uniform field with the same average chromaticity and luminance. Recall that this, somewhat surprising, conclusion holds only for the average results. Each individual observer deviates from this result in significant and systematic ways. The interesting result is that observer's seem to deviate from the mean in a variety of ways that average out to indicate no effect at all. Results for some individual observers are analyzed below.

Figs. 8-12 show the average results for four different observers along with error bars representing plus-and-minus one standard error of the mean (intra-observer variation). Figure 8 shows results for observer mdf1 (the author's first session). Observer mdf1 shows a clear increase in perceived brightness with contrast (note: this is as originally hypothesized by this observer). The simultaneous contrast results are consistent with the brightness matching results and suggest that an equivalent background model with nonlinear luminance integration would be appropriate.

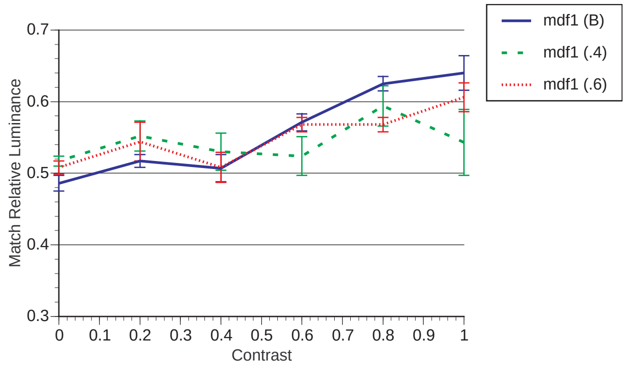


*Figure 8. Mean results for observer mdf1.*

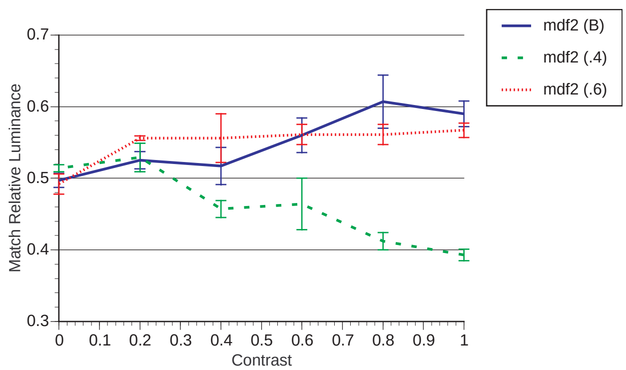


*Figure 9. Mean results for observer mdf2.*

Figure 9 shows the results for observer mdf2 (the author's second session). In this session, the observer adopted a different strategy in making the simultaneous contrast matches. For the mdf1 results, the observer matched the contrast of the patches with respect to the brightness of the backgrounds. For the mdf2 results, the observer matched the lightness of the patches with no direct concern for the background appearance (this is actually a more strict adherence to the instructions). The brightness matching results are virtually identical for mdf1 and mdf2. This is reassuring since the change in strategy should have no impact on the brightness matching results. The simultaneous contrast results for mdf2 show that the dark patch (0.4 relative luminance) looks lighter when the contrast of the background increases while the light patch (0.6 relative luminance) looks darker when background contrast increases. This is consistent with the results of Zaidi *et al.*[5] and suggests some form of contrast gain control. Thus, for observer mdf2, brightness increases with contrast (nonlinear integration) and contrast of image elements decreases (contrast gain control). There is no need to use the concept of atmosphere to explain this observer's results.

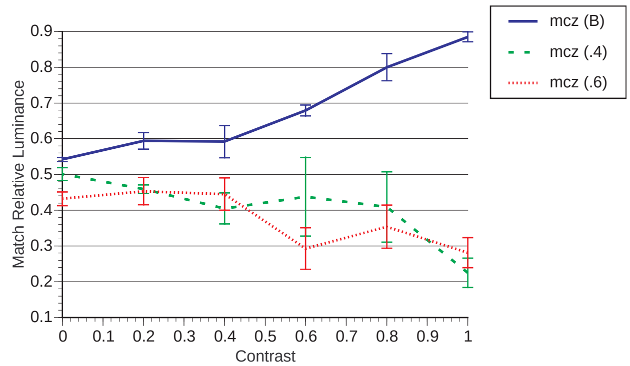


*Figure 10. Mean results for observer mcz.*

Figure 10 shows the results for observer mcz. Observer mcz showed the largest increase in brightness with contrast of any observer. This also suggests a nonlinear integration with an expansive function. Another way to phrase this interpretation is that the observer might have been keying in on one of the lighter patches to make his overall brightness judgement. The simultaneous contrast results for observer mcz show significant increases in the brightness of both patches with increasing contrast. This could be interpreted in terms of apparent atmosphere by assuming that the atmosphere is clearing as contrast increases (*i.e.*, less fog) and therefore the patches must be getting lighter (higher reflectance) in order to be of the same luminance. Observer mcz can be thought of as a prototype of the mean results without the effect being diluted by the wide variance of all the other observers.

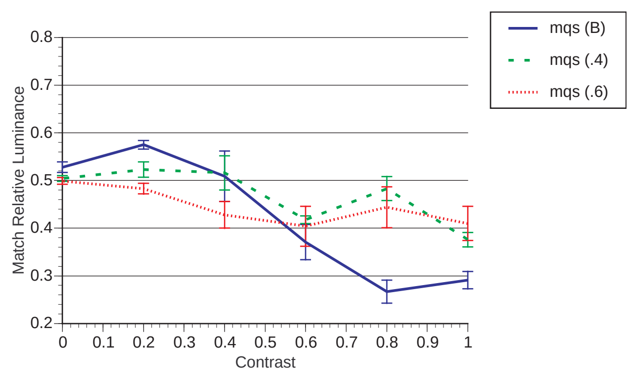


*Figure 11. Mean results for observer mqs.*

Observer mqs shows entirely different results as illustrated in Fig. 11. The simultaneous contrast results are similar in direction to those of observer mcz and the mean and thus can be interpreted similarly. However the brightness matching results are in the opposite direction. For observer mqs, brightness actually decreased with increasing image contrast at constant luminance. When observer mqs was interviewed about these results, he explained that as the contrast increased, it appeared that there was more black in the image and thus it was darker. It is interesting to note that observer mqs has a strong printing background and is accustomed to thinking of images in terms of density instead of lightness. Similar results were found in another observer with substantial printing experience. The results of observer mqs were confirmed in a brief follow-up experiment in which observers mqs and mdf alternated making brightness matches to make sure that they both were self-consistent and truly disagreed in such a profound manner. The results as illustrated in Figs. 8 and 11 were confirmed.

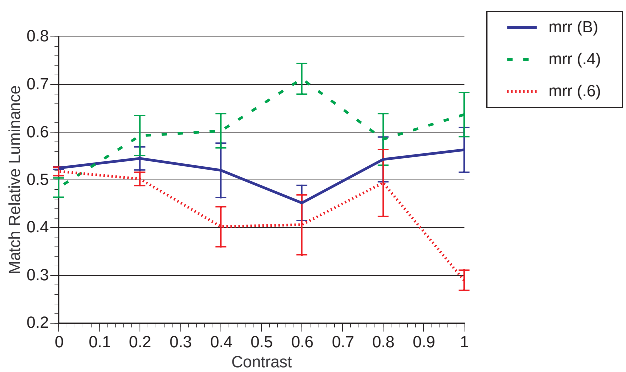


*Figure 12. Mean results for observer mrr.*

Lastly, the results for observer mrr are shown in Fig. 12. Observer mrr shows no significant effect for the brightness matching experiment despite expressing the belief that the images looked brighter as contrast increased. This observer therefore acted like a radiometer with linear integration for the brightness matching task. The simultaneous contrast results for observer mrr are exactly reversed from those of observer mdf2. In other words, for observer mrr, the light patch looked even lighter and the dark patch darker as background contrast increased. This is consistent with Adelson's[9] results and interpretation in terms of an atmospheric transfer function. These results are also similar to those found by Schirrilo and Shevell.[10]

## Conclusions

The wide variation in individual observers' results help to explain many previous results. It accentuates the point that the idea of the overall brightness of a scene is a high level perception that is not driven by low level sensory mechanisms. This is further confirmed by the fact that instructions, or observer strategy can impact the results. The results of various observers are consistent with previously published results that seem to contradict one another. This suggests not only that individual observer differences are important, but that small details in the stimulus configuration and task might have profound effects on the experimental results. Lastly, the fact that, over a fairly large group of observers, the results average out to indicate essentially no effect is fascinating. While each individual clearly sees an effect, the individual differences are such that the overall effect is nil – on average. This helps to explain why individual observers can be completely convinced that the predictions of a given color appearance model are incorrect while experiments for large groups of observers confirm the model's good performance on average. This result should also be a fair warning to take the results of visual experiments with a grain of salt when they involve higher-level perceptual mechanisms and only a few observers.

In conclusion, the results presented in this paper bode well for the use of color appearance models that treat spatial properties in a very simple way such as CIECAM97s. While there is certainly much to be gained with models that properly treat spatial properties of images,[11] on average models that assume linear integration and equivalent backgrounds should work quite well. Of course, individual results will vary.

## References

1. M.D. Fairchild, *Color Appearance Models*, Addison-Wesley, Reading, Mass. (1998).
2. OSA Colorimetry Committee, *The Science of Color*, Optical Society of America, Washington, (1963).
3. P. Oskoui and E. Pirrotta, Influence of Background Characteristics on Adapted White Points of CRTs, *IS&T/SID 6th Color Imaging Conference*, Scottsdale, 22-26 (1998).
4. R.O. Brown and D.I.A. MacLeod, Color Appearance Depends on the Variance of Surround Colors, *Current Biology* **7**, 844-849 (1997).
5. B. Spehar, J.S. DeBonet, and Q. Zaidi, Brightness Induction from Uniform and Complex Surrounds: A General Model, *Vision Res*. **36**, 1893-1906 (1996).
6. Q. Zaidi, J.S. DeBonet, and B. Spehar, Perceived Grey-Levels in Complex Configurations, *IS&T/SID 3rd Color Imaging Conference*, Scottsdale, 14-17 (1995).
7. J.S. DeBonet and Q. Zaidi, Comparison between Spatial Interactions in Perceived Contrast and Perceived Brightness, *Vision Res*. **37**, 1141-1155 (1997).
8. Q. Zaidi, B. Spehar, and J. DeBonet, Adaptation to Textured Chromatic Fields, *J. Opt. Soc. Am. A* **15**, 23-32 (1998).
9. E.H. Adelson, Lightness Perception and Lightness Illusions, in M. Gazzaniga, Ed., *The Cognitive Neurosciences, 2nd Ed.*, MIT Press, Cambridge, (in press).
10. J.A. Schirillo and S.K. Shevell, Brightness Contrast from Inhomogeneous Surrounds, *Vision Res*. **36**, 1783-1796 (1996).
11. S.N. Pattanaik, M.D. Fairchild, J.A. Ferwerda, and D.P. Greenberg, Multiscale Model of Adaptation, Spatial Vision, and Color Appearance, *IS&T/SID 6th Color Imaging Conference*, Scottsdale, 2-7 (1998).