ROBUSTNESS OF THE MUNSELL SCALE WITH REGARDS TO LUMINANCE CHANGES ON ITS BACKGROUND.

Olga Daneyko*, Tiziano Agostini*, Daniele Zavagno**
*Department of Psychology, University of Trieste, Trieste, Italy;
**Department of Psychology, University of Milano-Bicocca, Milano, Italy

Abstract

The constancy of a 16 step achromatic Munsell scale was tested with regards to luminance and hue variations in its background. Participants were asked to find the "colour" of targets on Munsell scales placed on four different backgrounds (each scale-background corresponded to a separate matching session): a white-black and a blue-red chequered background, and two homogenous grey backgrounds which luminance approximated the average luminance of the previous two backgrounds. Results do not support a background averaging hypothesis and indicate that the Munsell scale is affected by background luminance even when the background is articulated as a checkerboard.

In a previous study we tested the robustness of a 16-step achromatic Munsell scale with regards to background variations in two experiments (Daneyko, Zavagno, & Agostini, 2009). Participants were asked to find on a Munsell scale the colour of grey targets inside simultaneous lightness contrast displays (SLC). In experiment 1 we used a Munsell scale seen against a white, a black, or a white-black chequered background. The factor "scale background" determined significant differences in target matches. In experiment 2, the background of the Munsell scale consisted all in chequered backgrounds: white-black (from experiment 1), yellow-blue, and green-red. While the Michelson contrasts for the white-black and the yellow-blue backgrounds were fairly comparable, the Michelson contrast for the green-red background was significantly lower; nevertheless, the factor scale background did not determine significant effects. On one hand we have that the Munsell scale is not robust when background luminance values are extremely different for two scales otherwise equal; on the other hand we have that matching data from Munsell scales on different backgrounds are not statistically distinguishable when scale backgrounds are structurally highly articulated.

The present study is a continuation of the aforementioned research. In particular we looked into the results of experiment 2. Does the visual system average the luminance values of chequered backgrounds, ignoring even hue differences? Or a Munsell scale is simply robust when it is seen against highly articulated patterns, for instance, our chequered backgrounds?

The experiment

In the present study we used two chequered backgrounds and two homogeneous grey backgrounds which luminance approximated the average luminance of the two chequered backgrounds. The average luminance for a homogeneous grey background was calculated as the arithmetic mean of the luminance values of the corresponding chequered background. If the "background averaging" hypothesis holds, there should be no significant differences

between matching results obtained with a chequered background and a grey background which luminance is the mean of the chequered background.

Methods

Participants

Twenty-eight participants from the University of Milano-Bicocca (age 17-47, 18 female) volunteered to take part to the experiment. All participants had normal or corrected-to-normal vision and were naive to the purpose of the experiment.

Stimuli

The test stimulus were four achromatic Munsell scales, each made of 16 rectangular chips (6×2.2 cm) cut from Munsell papers ranging from 2.0 (black-left (52 cd/m^2) to 9.5 (white-right (1430 cd/m^2) (figure 1) and placed in a row (0.3 cm between each pair of chips). Each Munsell scale was placed on a different background (11.3×41.3 cm): a white-black chequered (w-b) background (checks measured 1.56×1.56 cm; the luminance of the white and of the black checks was 1290 and 270 cd/m^2 respectively), a red-blue chequered (r-b) background (280 and 255 cd/m^2 respectively), and two homogenous grey backgrounds (av.w-b; av.r-b) which luminance values approximated the average of the luminance values of the previous two backgrounds (respectively 730 and 290 cd/m^2). The steps of the scales were labelled in different ways: the scale on w-b showed the actual Munsell value labels; the scale on the av.w-b showed randomized numbers; the scales on the r-b and av.r-b backgrounds showed randomized letters.

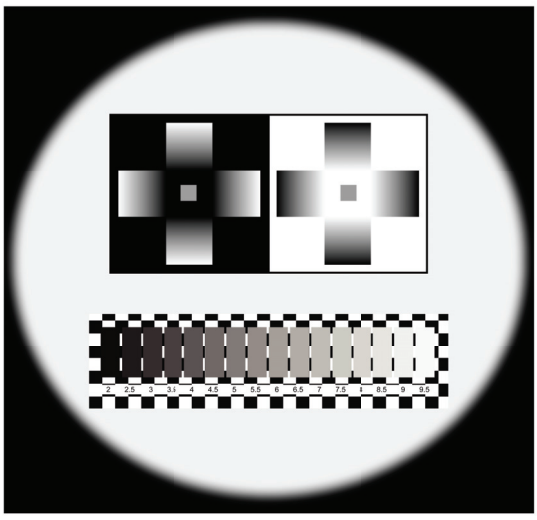

Figure 1. Example of the experimental set up with positive ramps and w-b scale.

Three comparison stimuli were used for the matching task: a classic simultaneous lightness contrast display (SLC) (figure 2a), and two modified SLC displays with luminance ramps: one showed the glare effect and the black hole effect (Agostini & Galmonte, 2002; Daneyko & Zavagno, 2008; Zavagno, 1999) (positive ramps, figure 2b); the other showed similar configurations but with inverted ramps (negative ramps, figure 2c; Daneyko & Zavagno, 2008). The black and the white backgrounds measured 20×20 cm and

had luminance values 83 and 2250 cd/m^2 respectively. The arms of the ramp crosses measured 6×6 cm and ranged from 83 to 2250 cd/m^2. The grey targets (2.2×2.2 cm) placed at the centre of each background were cut from Munsell paper 6.0 (R 30%, 790 cd/m^2). The corresponding 6.0 step on the Munsell scales had luminance 595 cd/m^2.

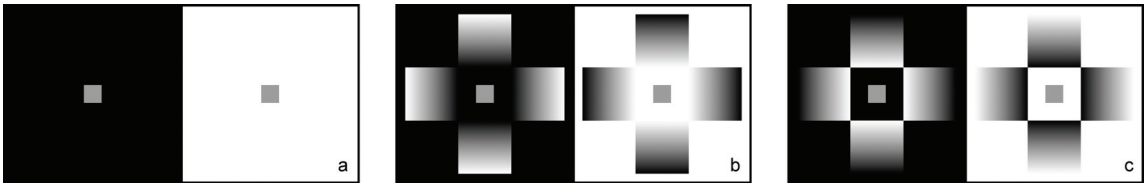

Figure 2. Comparison stimuli: (a) classic SLC; (b) SLC with positive ramps; (c) SLC with negative ramps.

The purpose of using different comparison stimuli is to introduce a nuisance variable (McBurney & White, 2007) for the psychology students who are usually familiar with SLC displays. In our previous experiments (Daneyko, Zavagno, & Agostini, 2009) we used instead two different luminance values for targets in otherwise classic SLC displays. Here we employed instead modified SLC displays, which produce significantly different lightness effects (Agostini & Galmonte, 2002; Daneyko & Zavagno, 2008).

Procedure

Each participant performed the matching tasks individually. Stimuli were viewed at a distance of 122 cm in a dark laboratory illuminated only by a theatrical spotlight, which shined on both the SLC display and the Munsell scale placed 14 cm below (figure 1). The spotlight was placed 42 cm behind and 70 cm above participant's sitting position. The task was to find the two chips on a Munsell scale that matched in colour the targets in the SLC display, as if a target and its corresponding scale chip were cut from the same grey paper. If participants could not find a perfect match they were asked to find the best match possible for each target.

A within group design was employed, meaning that each participant performed the matching task for all the comparison stimuli using the four Munsell scales in separate matching sessions. Comparison stimuli were presented in random order right side up or upside down. The matching task could be performed starting either with a target on the black background or with a target on the white background. After all matches were performed with one Munsell scale, the experimenter walked out of the participant's sight with all comparison stimuli and the scale just used, to return shortly after with another scale and the same three comparison stimuli (SLC displays).

Results

Experimental results are shown in Figure 3. An ANOVA for repeated measures (Scale-background×SLC-display×Target-background) was conducted on the data: all factors produced significant effects (respectively: $F_{3,\ 81}=2.92$, $p<0.01$; $F_{2,\ 54}=5.24$, $p<0.01$; $F_{1,\ 27}=148.75$, $p<0.0001$). Also the interaction SLC-display×Target-background and Scale-background×Target-background determined significant effects (respectively: $F_{2,\ 54}= 148.71$, $p<0.0001$; $F_{3,\ 81}=4.59$, $p<0.01$). While the first interaction confirms that our modified SLC displays determine different lightness effects with respect to the classic SLC display, the

second interaction suggests that the matching scale delivers different results in a selective way, depending on the comparison stimulus. To have a better look at this second issue for each target's mean match we ran a series of paired t-tests to compare the results obtained with Munsell scales on different backgrounds (p-values are reported in Table 1).

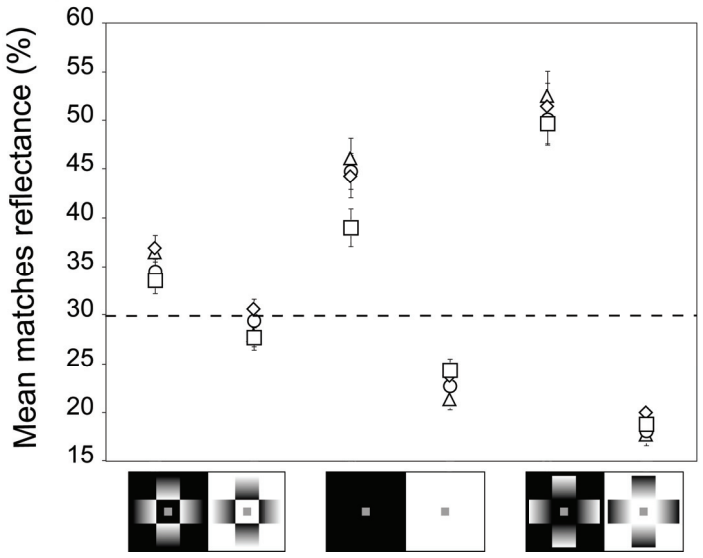

Figure 3. Result of the experiment. Symbols indicate the Munsell scale background: □ w-b chequered; ○ r-b chequered; ◇ grey av.w-b; △ grey av.r-b. The dashed line shows targets' actual reflectance.

Table 1: p-values for paired t-tests

	classic SLC					
Munsell scale backgrounds	**white backgrounds**			**black backgrounds**		
	grey av. w-b	*red-blue*	*grey av. r-b*	*grey av. w-b*	*red-blue*	*grey av. r-b*
white-black	p=0.6	p=0.7	p<0.05	p=0.06	p<0.005	p<0.05
grey av. w-b		p=0.3	p<0.05		p=0.7	p=0.3
red-blue			p=0.2			p=0.5
	SLC with positive ramps					
white-black	p=0.2	p=0.5	p=0.1	p=0.3	p=0.8	p=0.1
grey av. w-b		p<0.05	p<0.01		p=0.6	p=0.6
red-blue			p=0.6			p=0.3
	SLC with negative ramps					
white-black	p<0.05	p=0.1	p=0.8	p=0.06	p=0.5	p=0.2
grey av. w-b		p=0.4	p=0.059		p=0.1	p=0.7
red-blue			p=0.3			p=0.3

As one can see, the pattern of results is not simple. However, the graph in figure 3 and the results reported in Table 1 clearly show that matches obtained for same targets in classic SLC displays are affected by the background of the Munsell scales (MSs). In

particular, the mean match for the black background target obtained with the w-b chequered background is significantly different from the two means of the matches obtained with the r-b and the av.r-b backgrounds. However, there is also a tendency towards a significant difference between the w-b mean and av.w-b mean. The mean matches for the target on the white background appear instead more stable: differences appear only between the means for w-.b and av.r-b, and between the means for av.w-b and av.r-b.

Mean matches for targets in SLC displays with positive ramps are more stable. However, two important differences still emerged and they concern the white background target: the mean match obtained with av.w-b is statistically different from those obtained with r-b and av.r-b.

Mean matches for targets in SLC displays with negative ramps were also relatively stable. There is a statistical difference between the mean matches for the white background target obtained with w-b and av-w-b. A tendency towards a difference is also reported for matches for the black background target obtained with the same two scales.

Discussion

In general, significant matching differences were found mostly with classic SLC targets. If we compare these results with those of experiment 2 in our previous study (Daneyko et al., 2009), we can say two things: 1) data do not support the background averaging hypothesis; 2) background luminance can affect the appearance of the scale even when the background is articulated as a checkerboard.

The first point is actually somewhat tricky. In exp. 2 of our previous study the difference in average luminance between the backgrounds of the Munsell scale was about 80 cd/m^2; in exp. 1 of the same study the difference between the homogenous white background and the homogeneous black background was over 2000 cd/m^2, while the difference between those two backgrounds and the average luminance of the white-black chequered background was over 1000 cd/m^2; in the experiment reported here, the difference between the average luminance of the two chequered backgrounds is 440 cd/m^2. It could be that in this experiment we approach the limits of the luminance range within which Munsell scales are somewhat stable in their appearance. Hence the pattern of results is not uniform. More experimental research is required to clarify this point.

The second point comes from the fact that matches for black background SLC targets with the two chequered backgrounds are statistically different. This second fact, however, points to another issue: the scale appears robust for some targets and not for others. This can be seen also with the modified SLC displays.

Another relevant question is why the scale appears overall more robust when used in matching tasks with modified SLC displays. We believe that the reason is related directly to the lightness effects generated by luminance ramps applied to SLC displays. With positive ramps, the lightness difference between the two targets is considerably enhanced (Agostini & Galmonte, 2002; Daneyko & Zavagno, 2008) and probably pushed to its maximum: about 3 Munsell steps separate the two targets in SLC with positive ramps against the 1.5 Munsell steps for the classic SLC display. With negative ramps the lightness difference between the targets is considerably reduced (Daneyko & Zavagno, 2008), the lightness effect is at its minimum: only about half a Munsell step. In both cases we hypothesize that the effective "matching window" on the Munsell scale for the targets in the two modified displays (figure 2b-c) is greatly reduced: even if the scales appeared different, only very few choices were made available to carry out the matching tasks.

Finally, it is interesting to remark that lightness matches obtained with the scales seen against the red-blue chequered background and those obtained with the grey

average of the red-blue background are not statistically distinguishable, meaning that in our test stimuli the hue difference among scale backgrounds did not affect lightness matches.

Conclusions

The achromatic Munsell scale is a useful tool in lightness studies, for it allows the matching paradigm, which is both easy to set up and easy to analyse. Nevertheless, the Munsell scale is a lightness scale, and as such it can be subject to modifications in its appearance determined by changes in surrounding photo-geometric features. We addressed the robustness of the Munsell scale for what concerns luminance and hue variations in the scale's background. We found that scale background hue apparently does not affect lightness matches, while scale background luminance does. We also found that the effect of scale background luminance on lightness matches depends also on the appearance of the targets to be matched. Our next goal is therefore to conduct new experiments using other lightness effects as comparison stimuli in order to better understand the strengths and limitations of the Munsell scale as a lightness measurement tool.

Our experiment also wanted to test the possibility of a luminance averaging function carried out on the background of achromatic Munsell scales. Our data does not support the existence of such a function. Nevertheless, we believe that the issue is worthy of further empirical investigations carried out by manipulating other photo-geometric features, such as spatial frequency and background luminance articulation.

References

Agostini, T., & Galmonte, A. (2002). A new effect of luminance gradient on achromatic simultaneous contrast. *Psychonomic Bulletin & Review,* **9**, 264-269.

Daneyko, O. Zavagno, D Agostini, T (2009). Measuring the meter: On the robustness of the Munsell Neutral Value Scale with different surrounds. *Perception* **38** ECVP Abstract Supplement, 97.

Daneyko, O. & Zavagno, D. (2008). Simultaneous lightness contrast with non-adjacent luminance ramps and Gelb illumination. *Perception*, **37** ECVP Abstract Supplement, 104.

McBurney, D. H., & White T. L. (2007). *Research methods*. Belmont, CA: Thomson Wadsworth.

Zavagno D. (1999). Some new luminance-gradient effects. *Perception*, 28, 835–838.