Comparison of Unique Hue Stimuli Determined by Two Different Methods Using Munsell Color Chips

Renzo Shamey, Michael G. Sedito, Rolf G. Kuehni*

Department of Textile Engineering, Chemistry, and Science, North Carolina State University, Raleigh, NC 27695

Received 29 December 2008; revised 19 April 2009; accepted 9 May 2009

Abstract: Unique hue stimuli were determined by male and female observers using two different visual experimental procedures involving Munsell color chips of varying hue but identical chroma and value. The hypothesis was that unique hues can be more reliably established by explicit selection from a series of ordered stimuli than implicitly by hue scaling a series of stimuli in terms of neighboring UHs and this was statistically confirmed. The implicit selections based on long term memory of UHs appears to have been more challenging to observers since variability was increased by nearly 50% compared to when UHs were explicitly selected. The ranges of unique hues selected in the two methods were, however, comparable and no statistically significant difference was found between the results of females and males. The intra-observer variability in picking a stimulus to represent a unique hue, for all observers and averaged for all hues, was approximately 12% of the mean spread of unique hues, confirming that the large inter-observer variability is driven by differences in color vision and perhaps cognitive processes.

INTRODUCTION

According to Ewald Hering’s theory unique hues (UH) are the four fundamental perceived hues yellow, red, blue, and green, from which, in mixtures of neighboring pairs, all other hues are formed. Unique yellow is the yellow percept that to the observer appears neither greenish nor reddish, with a comparable situation in case of the other three UHs. The stimuli that cause perceptions of unique hues can vary considerably for color-normal observers, for as yet unknown reasons, as several investigations have shown.1 Although opponent mechanisms of cone signals have been located, for example in the lateral geniculate nuclei (LGN),2 their measured spectral responses are not directly related to unique hue percepts,3 indicating that if the latter have fundamental meaning their inputs involve individualized additional processing beyond the LGN.

Some of the earliest quantitative determinations of UH stimuli were in 1955 performed by Hurvich and Jameson using a hue-cancellation method involving spectral lights.4 The results are psychological data that fix, for a given observer, blue, green, and yellow UHs at specific wavelengths. For many observers unique red is located outside the spectrum and obtained by a mixture of lights from both ends of the spectrum.5,6 Since that time a number of additional UH determinations using spectral lights have been performed. Other methods are based on monitor displays and physical samples in form of color chips (for a summary up to 2004 see Ref. 1) or on other tasks.

There are several approaches to determine the stimuli that, for a given observer, result in the perception of UHs. Here, we are considering three of the basic methods as follows:

1. Create spectral functions by determining at applicable wavelengths the amount of spectral light seen as having a UH, say yellow, required to neutralize the perceived blueness in light of another wavelength, such as the blueness in violet. The resulting spectral functions are normalized and those of the opposing pairs arbitrarily expressed as positive and negative. UHs are located at the wavelengths where the function of the opposing pair passes through the line of neutrality.
A modified version of this procedure is to have the observer estimate the percentages of the two constituting UHs in the percept of a given wavelength, with the total in each case adding up to 100%.

2. Display lights of a given wavelength, monitor lights, or color chips singly and have the observer judge if the stimulus represents a UH.

3. Display ordered series of color chips and have the observer indicate those that represent UHs for them, usually by forced choice.

Past experiments by various authors have differed in regard to the state of adaptation of the observer, the surround, the length of the exposure to the stimulus and other factors, complicating comparison. A procedure used previously involves the display of 40 Munsell hue chips with identical lightness and chroma ordered sequentially on a neutral gray rotatable tray. Maximum chroma of complete hue circles of 40 samples at constant lightness (values 5 and 6) is limited to C8. A peculiarity of the Munsell system is that red hues at this chroma level appear chromatically less strong than other hues, a matter that sometimes has the advantage of not restricting the number of samples displayed for a particular hue, which avoids potentially biasing the observer with the choice of a restricted number of samples per hue. The results obtained by this method are well within the inter-observer variability of studies using more saturated stimuli.

In 2005, Abramov and Gordon reported on experiments using monochromatic lights as rear-projected stimuli against a black background, with a visual field size of 1° and exposure time of 0.5 s, with 15 s interval. Observers scaled the randomly displayed stimuli in terms of UH percentages. Each stimulus was shown four times, presumably in the same session. The agreement between observers was described as “very good,” and the wavelength range of the resulting personal UH stimuli varied from 28 to 83 nm with blue ranging from 427 to 483 nm (56 nm), green from 485 to 568 (83 nm) and yellow from 565 to 593 nm (28 nm). These three ranges essentially completely fill the spectral scale from 427 to 593 nm and are comparable or larger than those reported by most other investigators or the ranges of dominant wavelengths represented by the Munsell samples selected by method 3 described earlier.

We are not aware of reported studies investigating unique hue stimuli selections by observers according to method 3 above followed by the estimation of the percentages of the two constituting UHs in the percept of a given stimulus according to method 1, but using Munsell chips as reference samples to be scaled. Given the different cognitive tasks required in the two processes mentioned above it seems likely that the unrestricted pick of ordered samples from a rotating tray results in less intra-observer variability than the UHs resulting from scaling of intermediate hues.

**PURPOSE OF EXPERIMENT**

The purpose of the present experiment was to test the hypothesis that UHs can be more reliably established (less intra-observer variability in stimulus selection) by direct choice from a series of ordered stimuli than indirectly by hue scaling a series of stimuli in terms of neighboring UHs. For this purpose observers were asked to complete two tasks. In Task 1 the observer determined their UH stimuli by making simple choices from an ordered series of stimuli, being able to compare, say for red, the yellowness or blueness of several samples before arriving at a choice. Task 2, on the other hand, is a mental matching task of a series of color stimuli, with the ingredients consisting of UH choices from Task 1 or perhaps from mental concepts of unique hues, either stored in long-term memory or innately available to the observer by another process. The cognitive tasks in the two experiments are significantly different and a strong agreement would point to a very firm mental concept of UHs. A significant difference in reliability of determining UHs by the two methods would indicate that the UH concepts are less firmly stored in long-term memory and to a degree subject to influence from other mental processes.

A second purpose was to determine how widely the concept of UHs is known in a general student population, as indicated by the successful completion of Task 2.

**EXPERIMENTAL PROCEDURE**

A circular rotating tray was painted neutral gray (Munsell N 7.5) to match the color of the SpectraLight III (X-Rite) standard viewing booth. The illumination consisted of tungsten filtered D65 daylight. The 40 Munsell chips of the V6C8 hue circle were arranged in regular intervals near the periphery of the tray. Each chip was labeled and numbered in a random scheme. Observers were tested for normal color vision using the Neitz Test of Color Vision. Observers were required to wear gray gloves and laboratory coats to reduce variability in light reflected onto the samples. At the beginning of the experiment, in the otherwise subdued light in the room observers were asked to view the empty illuminated viewing booth for 2 min to adapt to the light source, during which time the experiment was explained. Two tasks were completed in each trial:

Task 1: The observer was asked to select the four chips from the series of 40 that to them represent the four UHs. The observer could rotate the tray at will and there was no time limit set. The coded information was recorded and the tray was moved out of the box.

Task 2: Every second chip (hue numbers 2.5 and 7.5) was sequentially shown to the observer in the viewing booth, in an order randomized separately for each observer. Observers were asked to estimate the percentages of unique hues in the hue of each sample, with the total adding up to 100%. No explicit instruction was given to limit the number of choices in each case to two.
UHs. The results in terms of UHs and percentages were recorded.

Each observer performed the test three times, with at least 24 h separation between trials.

Observers

A total of 30 observers with normal color vision according to the Neitz test (15 females and 15 males) participated. Most of the observers were students with little experience in color evaluation. Ages ranged from 18 to 50, with only two observers older than 30. All observers were acquainted with the concept of four UHs before the first test, without specifically pointing out that intermediate hues have perceptual components of only two UHs.

RESULTS

Task 1

Observers selected their UHs based on Munsell sample choices on a tray. The distribution of choices agrees well with distributions found in previous experiments of the same kind with more observers. Figure 1 shows the mean frequency of sample choices representing UHs from three trials for all observers.

The number of different hue samples selected as representing UHs is as follows, where the numbers in parentheses represent those from an experiment under comparable conditions with a total of 102 observers: Blue 6 (6), Red 7 (8), Yellow 5 (6), and Green 7 (8). There are no significant differences between the overall male and female responses.

The ranges of UH signatures of individual observers (the changes in hue angles in the Munsell hue diagram between UH choices) were also comparable with those from a previous study involving 102 observers and were as follows: uY to uG 45–102° (40.5–117°), uG-uB 63–123 (45–126), uB to uR 99–156° (81–148.5°), and uR to uY 39–99° (36–108°).

Task 2

In Task 2 (hue scaling), observers were asked to estimate the content of unique hues of 20 alternate chips from the 40 samples of the Munsell hue circle individually. For instance, 7.5R V6C8 Munsell chip was described as containing 80% red and 20% yellow. The task was complicated by the fact that the 20 samples were displayed randomly and without displaying observers’ previous choices of UH samples. Thus we believe the observers had to use their internal concept of unique hues as their guide to complete this task. Historically, three colors are considered to be primary in colorant mixture: yellow, red, and blue (or yellow, magenta and cyan). Hering’s perceptual fundamental colors, yellow, green, blue, and red, are much less widely known. As a result, even though all observers were acquainted with the concept of four UHs before the experiment, approximately one third of observers selected opposing UHs (particularly uY and uB) or three UHs as components of intermediate hues. In particular, the concept of uG as a primary was challenging for some observers as several used varying proportions of yellow and blue to identify green chips. Only 11 observers (six females and five males) always used neighboring UHs in all three trials to complete this task; additionally 7 observers (3 females and 4 males) used neighboring UHs in two of the three trials (interestingly, one of them in trials 1 and 2, one of them in trials 1 and 3 and five observers in trials 2 and 3). The remaining 12 observers used in some cases opposing UHs or three UHs to match the hues of the 20 samples or, in two cases, seemingly had conceptual problems with the task. The analysis of the results of Task 2 was therefore limited to the 18 observers that used neighboring UHs in at least two of the three trials. Figures 2(a) and 2(b) show two
typical examples of the scaling of the 20 hue samples, one that is remarkably consistent, given the random presentation of the samples, the other indicating use of opposing UHs (yellow and blue), as well as discontinuities (the latter not a reason for not using the data).

For the 18 observers included in the analysis the UH component percentage data of each sample straddling the neighboring UHs for a given observer and trial were interpolated, unless the sample was explicitly selected as a UH (100% value). Interpolation was based on the relative percentages of the choices in the adjacent pair. Where the interpolation did not fall exactly on a given hue of the Munsell 40 hue circle, the result was rounded to the nearest Munsell hue. This applies also to cases where two samples were determined to have the same UH, as in case of uB shown in Fig. 2(a).

Results of Task 1 and Task 2 Trials

The analysis of the data of the 18 observers consists of determining the intra-observer variability in the initial UH sample pick and in the UHs implied in Task 2. The means of intra-observer variability for three trials of Task 1 were compared with those of the weighted means of 3 and 2 trials, respectively, of Task 2. A change of one Munsell hue step in the selection of a UH sample was given a value of 1 and comparably for larger changes.

Table I shows that the mean change for the 18 observers is 0.86 hue step per UH from one trial to the next which is comparable to that (0.87) found in the previous study. The ranges indicate considerable differences in this respect between observers. Both investigations also show that the intra-observer mean variability represents only approximately 15% of the inter-observer spread of UH stimulus choices, indicating that there are relatively stable individualized concepts of UHs in the minds of the observers.

In Task 1 the observers make simple choices from stimuli present in front of them, being able to compare, say for red, the yellowness or blueness of several samples consecutively to arrive at a choice. Task 2 is a mental matching task with the ingredients consisting of mental concepts of unique hues. The emergence of the UHs seemingly depends on the long term memory of them. This appears to have been more challenging to observers since in Task 2 variability is increased by nearly 50% compared to Task 1, as shown in Table II. A comparison of mean change score per UH between observers completing Tasks 1 and 2 is shown in Fig. 3 which shows significant variation amongst observers.

The change scores between Task 1 and Task 2 have been compared for statistical significance of the differences and for correlation. For the 18 observers the p-value of a two tailed paired t-test is 0.016 indicating that the
difference is statistically significant at a 95% confidence limit. The correlation coefficient for the two sets of data is $-0.11$ indicating very low correlation between change scores for observers completing Tasks 1 and 2.

Comparing Task 1 and Task 2 Choices

Table III contains the results of the comparison between the individual results in the same test for the two tasks. The results are not directly comparable with those of Tables I and II because the number of opportunities for variation by observer in Tasks 1 and 2 is two per hue for the group of 11 observers completing three trials and one for the group of 7 observers completing two trials. In the comparison in Table III there are three sets of comparisons per hue for the group of 11 observers and 2 for the group of 7 observers.

The table indicates that the difference in the UH stimuli selected by an observer is on average slightly more than 1 Munsell hue step. Analysis of the direction of differences in selections did not disclose any discernible patterns of change both of individuals as well as in the groups. Figure 4 shows the range of Munsell chips selected as unique hues (filled sections) as well as the implicit unique hues (arrows) obtained via chip component analysis.

In the group of 11 observers completing three trials, out of 132 opportunities to have a difference between the initial UH pick and the subsequent implicit UH, in 27.3% of the cases there was no change. In this group the highest number of no-difference cases by observer was 6 (50%). There was one case of a 5-step jump. In the group of 7 observers, with 56 opportunities for change, 26.8% of the cases had no difference; the highest number of no-difference cases by observer was 5 (62.5%); there were 2 cases of 4-step jumps. It is evident that the interobserver variability for the 18 observers is relatively large. The observer with the lowest relative score had 7 step changes in 12 opportunities or 0.58 per opportunity; the one with the highest score had 26 step changes in 12 opportunities or 2.17 per opportunity, a factor of 3.7. The total results indicate significant variability among observers in terms of the mental UH concept.

### TABLE III. Intra-observer variability in difference between task 1 and task 2. Change score means and range.

<table>
<thead>
<tr>
<th></th>
<th>3 Trials (6F, 5M)</th>
<th>2 Trials (3F, 4M)</th>
<th>Mean (9F, 9M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>1.08</td>
<td>1.13</td>
<td>1.10</td>
</tr>
<tr>
<td>Males</td>
<td>1.47</td>
<td>1.44</td>
<td>1.45</td>
</tr>
<tr>
<td>Total</td>
<td>1.26</td>
<td>1.30</td>
<td>1.28</td>
</tr>
<tr>
<td>Range</td>
<td>0.58–2.17</td>
<td>0.63–2.13</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSIONS

The hypothesis that observers on average determine their UH stimuli with higher reliability when they can be selected from a range of stimuli than when they are determined from hue scaling judgments of the content of UHs in multiple samples has been statistically confirmed. The reason may be the difference in the cognitive tasks, in the former case having multiple stimuli simultaneously present that allow for a direct assessment. In the second, latter task, the result depends entirely on a mental concept of unique hue that may depend on long-term memory or some other neural apparatus. The total ranges of unique hues explicitly selected from a range of stimuli as well as the range of implicit unique hue selections based on judgments of the content of UHs in multiple samples were found, however, to be comparable with slight increases in the case of implicit UB and UR selections.

The concept of unique hues for the purposes of Task 2, despite a brief introduction at the beginning of the tests, was not sufficiently grasped by approximately one third of initial observers. Most inconsistencies involved the use of opposing UHs to match mixed hues or three UHs to match one test hue. Results from these observers were
excluded from the analysis of Task 2 and the comparison with Task 1.

No statistically significant difference was found between the results of females and males. However, there are significant differences in the performance of different observers, the source of which is not known.

The differences between observers in the choices of unique hue stimuli in Task 1, in terms of both choice by hue and hue signature, are in close agreement with results obtained for a larger observer panel in previous experiments using the same method. In this experiment the intra-observer variability in picking a stimulus to represent a unique hue, for all 30 observers and averaged for all hues, is 12.2% of the mean spread of unique hues, confirming that the inter-observer variability is driven by differences in color vision and perhaps cognitive processes.

ACKNOWLEDGMENTS

The authors thank Amanda Giggard, Brian Edwards, and Lina Cardenas for their help in the running of this project as well as all observers who took part in the study.