Will It Blend?

Paulo Duarte Esperança Garcia
Instituto Superior Técnico
Universidade de Lisboa
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
Email: paulo.garcia@tecnico.ulisboa.pt

***Abstract*—Over the last years, research has been made to ascertain if the blending of two or more colors can convey, in a efficient way, the information contained in two or more variables, using color blending techniques. Nonetheless, previous investigation has not come to an agreement.**

With our user study, we have found that CMYK is the color model that best resembles the users' expectations, while the orange, green and yellow color-blending results are the ones which generate shorter distances to between the reference colors and the ones indicated by the users. On the other hand, the CIE-L*C*h* Color Model is the one which is farther apart from the users' mental model of color. We have also detected that there is a mild indicator that it exists a difference between the responses from Female and Male users, but only in a few color blendings. All the data collected was developed through a user study with 259 users, which was supported by an online user studies platform called *BlendMe!*.

We gathered a set relevant implications of using color blending techniques, in the InfoVis field of research, besides providing a set of questions which remain unanswered and could turn out as an interesting source of future work, and a cleaned dataset containing the data collected from the user study which could be further analyzed by other researchers.

***Index Terms*—colors, models, spaces, blending, InfoVis, color perception, user study, calibration, scales, organization, data visualization, data processing, blending techniques, color vision deficiencies, demographic analysis**

I. Introduction

As it is stated by Chirimuuta [1], Color is a subjective interpretation of an objective physical *stimulus*, which may differ from person to person. We, as humans, do not equally perceive color: by saying this, it is affirmed that the definition and the interpretation of a colored *stimuli* can diverge depending on the philosophical mindset a person follows. Color perception can be influenced by cultural patterns and the environment in which we evolved as a specie. As it is known, the human visual system can only perceive light from a well defined wavelength range (from under 400 nanometers until 750 nanometers, approximately) and, consequently, determining the spectrum of colors which is human perceivable. The colors are combined and represented by Color Models (*e.g.* RGB, HSV or CMYK) which have their own color gamut: the combination of all these gamuts could be represent in a Color Space like CIE-XYZ, which maps all the human perceivable colors in a Chromaticity Diagram, frequently named as a Horseshoe Diagram given its shape.

Color is, nowadays, remarkably used as a powerful tool to convey information: it is used on statistical graphics, cartographical data, information visualization and developers are eager to use color in their interfaces to create a better User Experience - when accompanied of an appropriate *Color Scheme*. Particularly, when showing data variables on a graphic, it is commonly associated to each variable a color and relationships between them are concluded by observing it. Certainly, it would be useful to combine variables in the same graphic, using a technique of *Color Blending*, conveying exactly the same information but, from a Computer Graphics perspective, in an economical way.

This technique has not been widely exposed and investigated, but some interesting advances have been made yet. It has been researched if the blending of colors for data visualization [2] would be a proper technique to convey information, so as for Visualizing Social Personal Information [3]; on the other hand, it is also important to understand if users are able to perceive different amounts of blended colors [4].

Even though investigation has been done, there are flaws and situations raised from them which remain to be fully tested and understood. These led us to state our reseach goal as:

Study if color blendings can be detected by the users, while testing if it is easier for users to estimate the pair of colors that resulted in a particular given blend, or reciprocally, to estimate which blend will result from a given pair of colors.

We decided to conduct a user study to fulfill it and answer a small set of objective questions. We developed a platform which would be capable of dealing with laboratory and online data: this way, the overhead of analyzing data from two different environment would be reduced and data was concentrated in the same place. The study should be divided onto four important phases which will be addressed shortly in Section III.

It was created a set of color blendings based on the primitives from the RGB and CMYK Color Model: Red, Green, Blue, Cyan, Magenta and Yellow. This set was formed by blendings of two, which was presented to the user in two forms: by being given the result of the color blending to the user and asked for the basis, or by giving the basis and asking for the result which the user though it would be appropriate. These colors were blended following the HSV, RGB, CMYK, CIE-L*a*b* and CIE-L*C*h* Color Models, interpolating the colors from each pair accordingly.

In the end of the study, the analysis revealed that the color blending which constantly yielded the best results across all color models is the **mixing of Red-Yellow, to achieve Orange**, while the mixture which provided the worst results when evaluating the distance from the user's answers and the ideal answers, is the blending of **Green-Magenta, resulting in a Blue shade**; these results are consistent with the fact that the human color perception is conditioned by the amount of Cones present in our eyes.

When analyzing the answers by Color Model, we can observe that the CMYK Color Model is the one which presents the best results across both study environments, while the CIE-L*C*h* Color Model is the one which typically provided the worst results across all color blendings. We have also found that, for color blendings which involved the Red color there was, in fact, evidence that the users sort the colors when indicating the blending-basis, revealing some mental color organization.

The focus of this Master thesis was to conclude relevant implications of using color blending techniques, in the Information Visualization field of research, which are going to be discussed later on this document.

A. Contribution

This dissertation aims to provide useful inputs about how color blending can be correctly used in Information Visualizations, and according to the users' expectations, leading to the following contributions:

- A set of results which determine the answers to the aforementioned questions, obtained with online and laboratory users.
- A user studies platform called *BlendMe!*, which is composed of four test phases, helping the process of creating a user profile, testing the calibration and color vision deficiencies and collect user feedback.
- A compendium of guidelines on how to use color blending in Information Visualization, produced based on users' results, establishing aspects about color models and blendings which provide the best results.

II. Background

A. Color Perception

The Animal visual system is prepared to distinguish a wide range of green colors since we evolved as a species surrounded by green vegetation and knowing what to eat was a matter of life or death. The human visual system is adapted to detect sharpness and color with great precision and sensitivity during the day light and night. Hence, the light which reaches our human visual system is converted in the retina, which contains specialized cells - photoreceptors - that convert light energy into neural impulses, which are send to the brain.

There are two main types of photoreceptors: **cones** and **rods**, retinal cells that respond to light due to the absorption of photons in their proteins; cones are concentrated in the fovea. Particularly, the first ones can be classified in three different types: S (Small), M (Medium) and L (Large), according to the type of wavelength which they are sensitive to; the absence of one or more type of these cones causes a color vision deficiency.

B. Mental Models of Color

Mixing colors is a constant: learning how colors are disposed in a color wheel happens in our childhood and, in most of the case, a subtractive color model like the **RYB** is taught. As Gosset *et al.* state [5], the usage or learning of subtractive color spaces, in childhood, modifies the mental color model of each person. Typically, these models are quite different from the **RGB** model, which may confuse the observer, since these types of models constitute additive color spaces and are commonly used in visual displays.

Mental models of color are created individually by each user, and it can be influenced by many characteristics from his environment: Color is even perceptually different among different countries, continents, environments and genders: as it is known[1] [6], women can detect and describe with much more detail color than men, since the photoreceptors of men take a little longer wavelengths to perceive hue of a color. Also, it was investigated the cultural influence of each individual in his color perception: researchers have explored[2] the differences from the western color perception, and tribal color perception; the researchers presented a circle of squares with different shades of green, to the Himba tribe from Northern Namibia. Surprisingly, they were able to detect a larger number of shades of green than a western, non-colorblind person, which may occur because their environment do not manifest as much colors, and they need to detect different shades to hunt and pick up vegetation and fruits, which traditional western communities do not need to do.

C. Color Models & Spaces

Mixing the three primary color light channels to match any color is no longer an oddity and it constitutes the basics of *Colorimetry*: it is the science used to quantify and describe the human perception of color. We can describe color as the following equation [7]: $C = sS + mM + lL$, where C stands for the color to be matched, *S, M* and *L* are the primary light sources used to create the final color and that are detected in three types of cones, *s, m* and *l* represent the precise amounts of each primary lights.

The concepts of Color Spaces and Color Models are often confused but, in fact, they do not present the same idea (although they do use similar conceptions). Color models exist to mathematically conceive a description of color, in which color spaces will be based and present the equivalent colors, while the latter represents the gamut of colors described accordingly to the primitives of a color model, containing

[1] "Where Man See White (...)", Available at: www.bit.ly/1AMHgcW. Last accessed on October 17th, 2016.

[2] "It's Not Easy Seeing Green". Available at: www.nyti.ms/1S71yVo. Last accessed on October 17th, 2016.

not only visible colors but also colors that are impossible to represent on physical devices.

Regarding Color Models, there are two types: **Additive Color Models** use light to display color mixing, mixing primary colors such as Red, Green and Blue; equally combined and overlapped, they form white light, whereas **Subtractive Color Models** mix colors using paint pigments and the result of any blend is a color that tend to be darker, the more you mix it. Examples of Color Models are RGB (Red-Green-Blue), HSV (Hue-Sauration-Value) and CMYK (Cyan-Magenta-Yellow-Key).

Color Spaces allow the representation of reproducible colors on a given physical device, while relating the description of a color model to actual colors, being a three dimensional object that contains all realizable color combinations. There are three types of them: **Device-Dependent**, **Device-Independent** and **Working Spaces**. The most important Color Space is the **CIE 1931** which, nowadays, is derived by **CIE 1931 XYZ** and **CIE 1931 RGB**; the first one is the most important, comprising all color sensations which a human can perceive, standing out as a standard for other color spaces. Another frequently used color space is the **CIE-L*a*b***, where two axis are represented by the a* and b*, being the first one representative of Red-Green and the latter the Blue-Yellow; the L* variable represents the lightness. This color space derives in a cubic color space representation, which is recognized as the **CIE-L*C*h***.

D. Related Work

Usually, when it comes to encoding information, color appears as the number one choice, due to the its ease of perception and familiarity. When representing more than one colored variable at the same time, it would be useful to perceive interrelations among them and if the users are able to understand which of these entities are related, or blended.

Gama and Gonçalves started their research [2] aiming to study to which extent people are able to, given a specific color resulting from a mixture of two colors, understand the blended color's origin; besides, they studied which is the color model that yields the most accurate results: hardware-oriented color models like RGB or color-printing models such as CMYK, fail to provide a color perception description, unlike HSV. This pitfall is amended by CIE-L*C*h*, by creating a perceptually uniform scale to lightness. On the other hand, when these researchers have tested the perception of relative amounts of colors [4], they have concluded that users happen to perceive most colors correctly regarding the pair Red-Yellow and, likewise, colors in both extremes, even for other pairs: "central colors" are generally the most problematic.

There was some research regarding the usage of color in InfoVis: particularly, about colormaps which are commonly used in computer science. Zhou and Hansen [8] provide a way to classify colormapping techniques into a taxonomy for readers to quickly identify the appropriate techniques they might use, classifying representative visualization techniques that explicitly discuss the use of colormaps; these authors gathered the investigation performed by other researchers in this paper. Additionaly, the authors classify colormap generation as: **procedural, user-study based, rule based** and **data-driven**. One good example of Color Blending applied with colormaps is the one studied by Stoffel *et al.* [9], since they introduced a new technique for visualizing proportions in categorical data; in particular, they combine bipolar colormaps with an adapted double-rendering of polygons to simultaneously visually represent the first two categories and the spatial location.

III. BlendMe! - The User Study

A. Objectives

The goals which we have picked for this User Study were to understand if color blendings can be detected by users, is it easier for users to estimate the pair of colors that resulted in a particular given blend, or reciprocally, to estimate which blend will result from a given pair of colors, to detect if the users follow some kind of mental convention and organize the color when conveying the answers, and formulate possible implications of color blending usage, in Information Visualization field of research.

We have planned to develop this study in two different strands: in a **Laboratory Environment**, which will allow us to calibrate and perfectly control the entire study conditions, and in an **Online Environment**, which will allow us to disseminate our study to a larger set of users. Therefore, we have defined the following questions:

- **Q1:** Which Color Model best meets the users' expectations, when blending two colors?
- **Q2:** Do users specify the Blending-basis following some order, when users are indicating possible color mixtures' results?
- **Q3:** Are there evidences from differences across demographic groups, such as the age or gender?

We drafted our study into four different phases: a **User Profiling Phase**, a **Calibration Phase**, a **Color Deficiency Test Phase** and finally, the **Core Phase**. In the following section, we detail each of these study phases.

B. Design

We intended to develop a user study which could support the laboratory controlled environment, while at the same time supporting the collection of metrics and data from the online users: this is an important consideration, since the workload when analyzing the results would be dramatically reduced because the data is condensed and gathered in the same fashion, and data would be more comparable. When brainstorming the ideas for this study, we started with the intention of testing both the blending of two colors and three colors; we decided that the colors which would be blended were Red (**R**), Green (**G**), Blue (**B**), Cyan (**C**), Magenta (**M**) and Yellow (**Y**), since they represent each primitive of the most commonly known Color Models, RGB (Additive Color Model) and CMYK (Subtractive Color Model). The color models we intended to study: the color models were **HSV, RGB, CMYK, CIE-L*a*b*** and **CIE-L*C*h***.

Then, we produced a wide spreadsheet of possible blendings of these colors, according to these color models, **mixed in pairs of two colors**: this generated the total amount of 78 blendings, **There are 15 possible mixtures of two colors**, when combining the previous defined colors: R-G, R-B, G-B, R-C, R-M, R-Y, C-M, M-Y, G-C, G-M, G-Y, B-C, B-M, B-Y and C-Y. Since we wanted to **test if it is better to give the result already mixed, or indicate the blending-basis and the users create the color mixture**, we developed a set of 32 questions to present to the user: 17 of them are of the type **presenting the resulting color, and ask for the blending-basis**, and 15 of them are of type **given the blending-basis, ask for the blending-result**. The entire set of questions is mapped in Table I.

Question ID	Given the Result, Asked for Basis Given Color	Question ID	Given the Basis, Asked for Result Given Colors	
1	#FFFF00	18	#FF0000	#00FF00
2	#FF00FF	19	#FF0000	#0000FF
3	#80FF00	20	#00FF00	#0000FF
4	#7F00FF	21	#FF0000	#00FFFF
5	#FF0080	22	#FF0000	#FF00FF
6	#FF8000	23	#FF0000	#FFFF00
7	#0000FF	24	#00FFFF	#FF00FF
8	#FF0000	25	#FF00FF	#FFFF00
9	#00FF80	26	#00FF00	#00FFFF
10	#0080FF	27	#00FF00	#FF00FF
11	#FF8000	28	#00FF00	#FFFF00
12	#80FF00	29	#0000FF	#00FFFF
13	#0080FF	30	#0000FF	#FF00FF
14	#8000FF	31	#0000FF	#FFFF00
15	#00FF80	32	#00FFFF	#FFFF00
16	#FF007F			
17	#00FF00			

TABLE I
TWO TYPES OF QUESTIONS ASKED ABOUT COLOR BLENDINGS

1) User Profiling: In these phase, questions were asked about the Age, Gender, Academic Degree, Nationality and Country of Residence: these questions helped us conceiving user profiles with key indicators about cultural background and gender relation to results of each test.

2) Testing Calibration: We developed another solution for remote controlling the calibration on the online environment: to present two similar calibration images, one presenting a set of shaded squares ranging from grey to black shades against a black background, and another presenting instead white squares against a white background. The iser had to indicate the number and word from the last square which he could easily see. This information provide us input about the white-level and black-level of the screen, which are nothing more than the **Contrast** and textbfBrightness, respectively, of the display. Regarding the laboratory environment, we conducted the users tests in a LCD monitor which was calibrated using a Spyder[3] Colorimeter, that will consider the existing light in the environment and adjust the color of each pixel to a standard.

3) Testing Color Vision Deficiencies: The Color Deficiency Test was comprised of a set of six plates, which were able to detect which type of color vision deficiencies the user would eventually have. This test in commonly known as the *Ishihara Test*, which has a validated [10] short form that rearranges the order in which the plates are presented. We have only chosen plates which detect color vision deficiencies in the Red-Green field, since it is the most common deficiency. The plates chosen were:

- Plate #1 - Presents the number **12**. Every user should be capable of viewing the same number.
- Plate #2 - Presents the number **29** for regular users which do not have any color vision deficiency, and the number **70** to users which have a deficiency in the Red-Green field.
- Plate #3 - This plate is a confirmation from the result of the previous one, presenting **74** to regular users and **21** to users which have a deficiency in the Red-Green field.
- Plate #4 - Presents the number **45** for the regular users, and the ones which have a color vision deficiencies are supposed to see a blob.
- Plate #5 - Presents the number **26** for the regular users, the ones which have Deuteranopia see only the number **2** and the ones which have Protanopia see number **6**.
- Plate #6 - Presents the number **35** for the regular users, the ones which have Deuteranopia see only the number **3** and the ones which have Protanopia see number **5**.

4) Core Test: This phase is the principal part of the study, which will evaluate the **Blending of Two Colors**. We have composed an interface with a small set of objects which would be used and interacted with to provide the colors to the user, and receive his input values, among other objects. We wanted to provide a tool which would be capable of displaying without being influenced by its surrounding of even by the proximity of other colors: then the colors were presented in rounded shapes, accompanied by what we call ***color sliders*** whenever it was needed input colors from the user. With this, only the necessary colors are displayed on the circles as the users wish and there is no interference of undesired colors, allowing us to eliminate the influence of them.

These shapes start filled with an empty color (or white) so it **does not influence color perception**, and the users are not influenced by previously used colors, when answering to another question. There is no particular reason for the chosen shapes are circles, since **we are not studying the best visualizations to convey information, when using color blending techniques**.

The most interesting fact of this test phase is how the color slider works: we chose the HSV Color Model to represent the colors to show, since the **HSV Color Model has the best compromise when presenting colors in information visualization** because of its primitives (Hue, Saturation and Value). **We have chosen to only modify the Hue value and leave the Saturation and Value on its maximum value**: this way, we could ensure that we could present the entire range of colors at its full saturation and value, also simplifying the gathering of values from the users. Therefore, the color slider yields a value within the range of $[0^o; 360^o]$ degrees which corresponds to an angle in the Hue circle of the HSV Color Model. Moreover, the scale of values of the slider, though representing continuous angle values, does not presents

[3]"Spyder - Datacolor Imaging Solutions", Available at: http://spyder.datacolor.com. Last accessed on October 17th, 2016.

the values ordered from 0º to 360º. Instead, **fixed intervals of values are mixed within each other**, so the users do not formulate any mental organization in the moment and do not demonstrate any previous conception or mindset. A representation of these color sliders in present in Figure 1, which depicts a wireframe version of these objects.

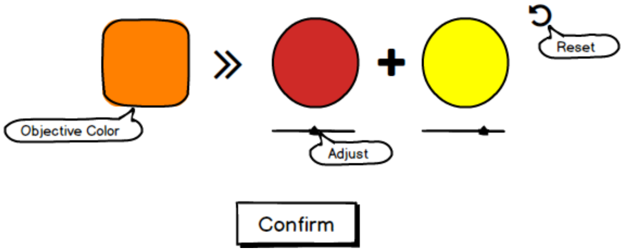

Fig. 1. Wireframe of Color Sliders

The colors represented in the shapes and the ones present in the pre-calculated answers have a particularity, in the laboratory environment: these pre-calculated values are, with the help of a *Matlab* script, going to be **converted to adapted color values, according to the .icc color profile file generated by the Spyder Colorimeter**. This way, we guarantee that no matter if the laboratory environmental conditions change, the colors will always be presented equally to every user. **This script processes each possible result for every color pair, in every color model**, converts it to a normalized CIE-XYZ value and in the end, to an hexadecimal color code to be printed on the shapes of the interface, when time comes for the laboratory user to choose a color from the color slider. This process was realized before each user session.

After the users indicated and confirms their answer, **they were presented a satisfaction question with a 5-point Likert Scale** to double-check the easiness of each mixture, as Gama and Gonçalves did [2].

IV. Analyzing Color Mixing Perception

The users were given always the same briefing when they arrived at the user study test site: it was explained the motivation behind the master thesis, the goals which were expected for this phase of the study and what was expected for them to execute. The user was told that "there are no pre-defined correct and wrong answers to each question, this test was designed to test the general color mixing capabilities of the majority of the users". The user study was self-contained, in the sense that every other relevant information and instruction was in the interface, adapted for each test phase, so it was not given any physical artifacts giving instructions.

In order to collect a larger amount of users (either for the laboratory or online study), we had to spread the study both by *word-of-mouth* and across some online platforms. We have opted for spreading the study across social networks like Facebook® or StumbleUpon. However, we found that gathering the laboratory users was a tough job to do, since many prospective users were not willing to fulfill the study. Thus, we explored a *Reddit* subthread called "Sample Size", which exists to disseminate user studies across the internet, which provided 159 new visitors to our webpage.

We collected a **total amount of 477 users** which interacted with our study and fulfilled, at least, until the Color Vision Deficiencies Test Phase. However, only **259 users went on to the core phase of the study**, representing **54.29%** of the total amount, giving at least one answer on the set of 32 questions; the 218 users that did not leave an answer amount to **45.71%**. There were **28 users who performed the entire study on the laboratory trials**. On the other hand, there were **231 users which carried out the study online**. The data presented and used in this dissertation document was gathered during roughly two months, from 15th of April until 8th of June.

Note that the main data to be analyzed is the one obtained in the laboratory environment, being the online data only the corroboration of the main data. Each chromaticity diagram presented in this Section, shows a **black filled dot** which corresponds to the ideal answer for each color model, and **two black empty dots** which are the blending-basis of questions which have asked for it; the **grey dots** represent the answers given by the users, and the black lines represent the union between each grey dot, forming each answer pair given.

A. User Profile

Our user sample is composed by 259 (100%) users, being **105 (40.5%) Females**, **152 (58.7%) Males** and a minority of **2 (0.8%) Other gendered users**: this sample age can be characterized as being generally young ($\overline{x} = 29.77$, $\overline{x} = 23$, $s = 40.30$), surprisingly having **8 users (3.09%) aged above 60 years old** which could enhance some interesting differences between age groups. Generally, our users have high academic qualifications, representing **66.02% of all users (Bachelor, Master and Doctoral Degrees)**, being **38 (14.67%) users qualified with College degree**, **47 (18.15%) have a High-School degree** and only **3 (1.16%) subjects do not presented any academic degree**. Between the laboratory and online environment, the distribution of users remains with the same proportions: more male users than females, mostly aged between 20 and 29 years old (60.71%) and the majority having a superior academic degree (46.43% BSc and 35.71% MSc). From the entire user set, **215 (83.01%) of them have Portuguese nationality**, **217 (83.78%) live in Portugal** currently and **216 (83.40%) speak Portuguese**; the second most influent group of users are the ones from english-speaking countries (United Kingdom, United States of America, and others). Other minor users which contributed to our survey came from Turkey, France, New Zealand, Sweden or even Antarctica (among others) - **these countries represent only 5.41%**.

B. Color Models

Since the colors obtained in the color slider from the Core Phase indicate values for the HSV Color Model, the values were converted from HSV to CIE-XYZ Color Model. Thus, we can produce color blends in every studied color model (HSV, RGB, CMYK, CIE-L*a*b* and CIE L*C*h*) and ensure that colors obey to the same common standard; also, this is specially important to produce Chromaticity Diagrams where colors are mapped according to a set of XYZ primitives. Afterwards, every resulting blending is compared to the pre-calculated value for each color model, and the euclidean distance to the latter value is stored for statistical analysis. It was this distance which we had considered when analyzing the relationship between the users' expectations and each color model.

The Color Model which presented the lowest mean value for distances was the CMYK Color Model ($\overline{x} = 0.10$), followed by HSV, RGB and CIE-L*a*b* (all with $\overline{x} = 0.14$), being CIE-L*C*h the model which has the highest calculated distances ($\overline{x} = 0.20$). Consistently, the color model which has the lowest variance and range of answers is the CMYK ($s^2 = 0.001$, $range = 0.14$), opposed to the HSV ($s^2 = 0.006$, $range = 0.26$). Performing a *Wilcoxon* Test, we conclude that there are no statistically significant differences between the laboratory distances and online ones ($p < 0.05$), which means that the online results corroborate the laboratory ones.

1) HSV Color Model: The questions which have shorter distances are number 11 (given orange and expected green and magenta), number 17 (given green, expected blue and yellow) and number 6 (given orange, expected red and yellow), with mean values of $\overline{x} = 0.0587$, $\overline{x} = 0.0683$ and $\overline{x} = 0.0741$, respectively. On the other hand, we found that the questions which generate the worst results are number 2 (given magenta, expected red and blue), number 13 (given a shade of blue, expected blue and cyan) and number 10 (given a shade of blue, expected green and magenta), with correspondent mean values $\overline{x}_2 = 0.2173$, $\overline{x}_{13} = 0.2713$ and $\overline{x}_{10} = 0.3$.

When comparing the results from the HSV color model with other color models' results, we detected some statistically significant differences with other models: performing a *Wilcoxon* Test ($p < 0.05$), we can infer that HSV does have statistically significant differences with CIE-L*C*h*, CMYK, RGB and CIE-L*a*b* in the majority of questions. Due to the fact that there are color primitives which are opposed in the HSV hue circle, there are in this study three pairs of colors which blend onto two possible colors: pairs Red-Cyan, Green-Magenta and Blue-Yellow. By placing these questions, we intended to understand which side the user would tend to follow when asked to indicate the blending-basis. In this article, we analyze just one of these.

The diagram of Figure 2 contains the three top and bottom-valued questions, disposed on top of an interval $[0; 0.5]$ of differences. Each question is mapped according to its mean value for distance to the ideal HSV Color Model response, while is accompanied by the range of values which compose its answers.

*2) CIE-L*C*h* Color Model:* This model has the highest mean value on laboratory and online environment, and the highest standard deviation, range and variance exclusively on the online environment. The three questions which have shorter distances are number 6 (given orange, expected red and yellow), number 5 (given a shade of red, expected red and magenta) and number 2 (given magenta, expected red and blue). The mean values distances for this questions were, contrary to the previous color model, are all values above 0.1 (a quite high value in the scale which results are presented), that could represent a significant change in the resultant color.

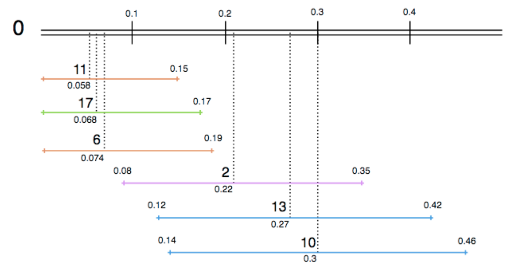

Fig. 2. Best and Worst Questions, according to HSV Color Model.

The model which CIE-L*C*h* has most statistically significant differences with is CMYK, with eleven questions, according to the *Wilcoxon* Test ($p < 0.05$). Question 6 presented again a constant top value for the mean value, on this model, which leads us to formulate the theory that the **orange color is commonly used and mixed by the users**. On the other hand, question 10 kept having one of the worst mean value ($\overline{x}_{10} = 0.2557$), strengthening the theory that **blue shades and tones will probably wild worst results, due to the fact that it is the color which human have less descriptive power**.

The diagram of Figure 3 contains the three top and bottom-valued questions, disposed on top of an interval $[0; 0.5]$ of differences.

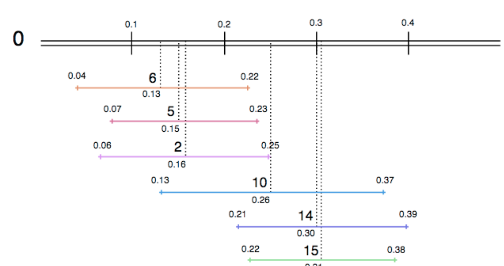

Fig. 3. Best and Worst Questions, according to CIE-L*C*h* Color Model.

3) CMYK Color Model: This model is, by far, the one which presents the best results: it has the lowest mean value of distances ($\overline{x}_{lab} = 0.10$, $\overline{x}_{online} = 0.09$), lowest standard deviation ($s_{lab} = 0.04$, $s_{online} = 0.04$), the lowest range of values ($range_{lab} = 0.14$, $range_{online} = 0.11$) and, finally, the lowest variance value ($s^2_{lab} = 0.001$, $s^2_{online} = 0.001$). The three questions which have shorter distances are number 17 (given a green, expected cyan and blue), number 6 (given orange, expected red and yellow) and number fifteen (given a shade of green, expected blue and yellow). The mean values distances for this questions were $\overline{x}_{17} = 0.0452$, $\overline{x}_6 = 0.0577$ and $\overline{x}_{15} = 0.0625$ which, contrary to the previous color model, are all values below 0.1, that does not represent a significant change in the resultant color.

Question 17 presented consistently another low value: along with question 3 and 15, these presented different shades of green and all had yielded favorable results. This may lead us to form another conclusion: **questions which presented shades of green color may produce better results, according to users' expectations**. This is consistent with the fact that **humans have more descriptive power in the green zone of the spectrum, due to the amount of cones which exist in the human eye**.

The questions which yielded **better results are majorly related to primitives of the CMYK color model**. This success could be related to the fact that **people tend to formulate mental models of color based on ink mixing in childhood** [5], mostly associating it to CMYK Color Model without even knowing it.

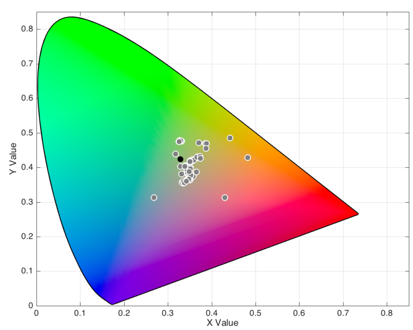

Fig. 4. Online: Question 17, Regular users, mixed in CMYK.

Figure 4 show the results for Question 17, the question with closer distances. The diagram of Figure 5 contains the three top and bottom-valued questions, disposed on top of an interval $[0; 0.5]$ of differences.

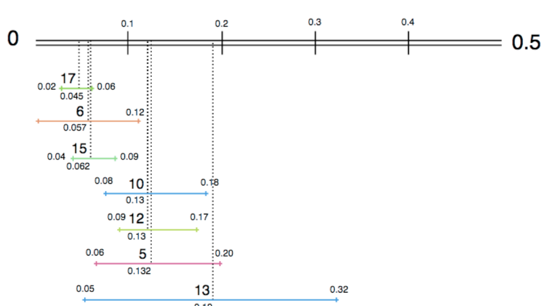

Fig. 5. Best and Worst Questions, according to CMYK Color Model.

4) RGB Color Model: This color model is complementary to the CMYK color model. Feedback collected from the users was such that, sometimes, the users which knew how to blend in subtractive color models, tended to be confused and tried to mix additive color models, also. Based on the results collected from the laboratory users, we can tell that the results from this color model are quite similar to CMYK results: RGB has one of the lowest mean value of distances ($\overline{x}_{lab} = 0.14$, $\overline{x}_{online} = 0.12$) and one of the lowest standard deviations ($s_{lab} = 0.05$, $s_{online} = 0.04$ equal to $s_{CMYK-online}$). The three questions which have shorter distances are number number 6 (given orange, expected red and yellow), number 9 (given a shade of green, expected green and cyan) and 17 (given a green, expected cyan and blue). The mean values distances for this questions were $\overline{x}_6 = 0.0777$, $\overline{x}_9 = 0.1005$ and $\overline{x}_{17} = 0.1030$.

When comparing the results from the RGB color model, the majority of users did reveal lack of knowledge in mixing the colors according to an additive color model, as they tended to mix colors according to the CMYK color model. However, this color model has a high degree of compatibility with every other model, excepting CMYK: performing a Wilcoxon Test ($p < 0.05$), we can infer that RGB does not present statistically significant differences with HSV, CIE-L*a*b* and with CIE-L*C*h* in 10 questions, while presenting statistically significant differences with CMYK in 13 questions.

Relating to the fact that people either tend to blend colors in an additive or subtractive way, we can compare the results from this color model with CMYK model and state that **users tend to formulate subtractive mental models of color blending**. However, **there is room for further investigation, to fully understand if users are influenced by additive color models or subtractive ones**. The diagram of Figure 6 contains the three top and bottom-valued questions.

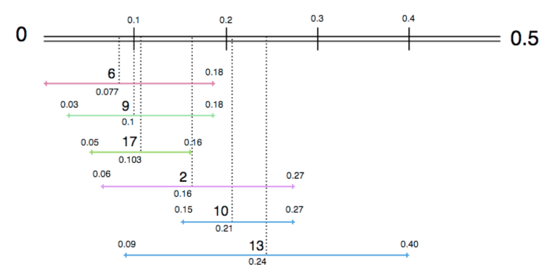

Fig. 6. Best and Worst Questions, according to RGB Color Model.

*5) CIE-L*a*b* Color Model:* This color model has results very similar to CMYK and RGB, since its descriptive statistics have comparable values to this model. The three questions which have shorter distances are number number 6 (given orange, expected red and yellow), 17 (given a green, expected cyan and blue) and number 9 (given a shade of green, expected green and cyan), which are exactly the same as before. The mean values distances for this questions were $\overline{x}_6 = 0.0509$, $\overline{x}_{17} = 0.1130$ and $\overline{x}_9 = 0.1142$.

Since the CIE-L*a*b* conveys the entire set of perceived by the human eye, it **explains why the results associated with this color model are so close to others which yield the best results**. The diagram of Figure 7 contains the three top and bottom-valued questions.

C. Color Blending Expectation

Besides asking our users to indicate us the two primitives which composed a color blending of two colors, we intended

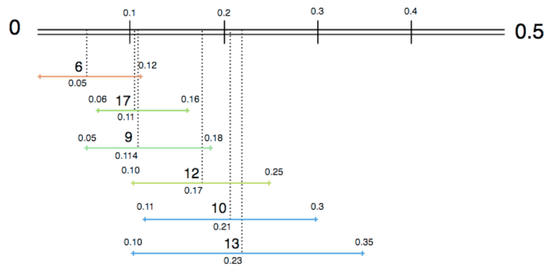

Fig. 7. Best and Worst Questions, according to CIE-L*a*b* Color Model.

to go further: **comprehend if more than detecting two colors of a mixture, a user is capable of mentally blend two given colors and indicate us its results**.

We end up concluding that the question which constantly had best results among all color models was the blending of red and yellow to create orange. However, these results represent much higher distances to the ideal answer in any color model, than when given the user the resulting color of the blending. This fact could accuse that **it is harder for the users to detect the result of a color blend, when two blending basis are given, than when the resulting color is given**, even for the orange color blending which produced very good results in the previous analysis.

Contrary to questions in which the user was asked to indicate the blending basis, **the color model which has the worst results is the HSV Color Model** ($\overline{x}_{HSV} = 0.18$) while the one which has **the shortest mean distance value continues to be CMYK, along with CIE-L*a*b*** ($\overline{x} = 0.13$).

D. Color Blendings & Naming

Since the creation of an entire study to ascertain this subject was out of the scope of this master thesis, we used the Color Survey[4] which the web page XKCD conducted, to study which were the most common RGB color triples among users. They performed roughly more than 222 000 user sessions to ascertain color naming, producing a map which shows the dominant names attributed to RGB colors over the faces of RGB cube; they also produced a file comprised of 196 608 RGB triplets, grouped by **Color Bins**. We decided to use it given its large sample size, and it was executed with a great amount of users which can verify it. We also found great compatibility between this survey and ours, since the values in the first were also presented in its maximum value of saturation.

The idea was to compare our answers with each color bin, creating a mapping between our users' values and commonly-used names; to speed up the computation of the comparisons, each color bin was drawn and the smallest polygon formed by the set of triplets of each bin was used to compare the values. There were six questions which presented as the result of a color mixture, either one of the following primary colors; what we intended to know is if the users are capable of creating color blendings which result in these colors. For the sake of simplification, in this document we will only present the analysis of creating Red based on Magenta and Yellow, and Green based on Cyan and Yellow.

[4]Color Survey Results. Available at: blog.xkcd.com/2010/05/03/color-survey-results/. Last accessed on October 17th, 2016.

1) Magenta + Yellow = Red: The laboratory users did focus their answers on **orange, green, blue and even red hue**; moreover, these results were similar among the online users and did not demonstrated a consensual answer pair. Additionally, five laboratory users (from thirteen which indicated a pair of two answers - 38.46% of the users) and forty-two online users (from sixty-two - 67.74% of the users) indicated, at least, one time the red color as an answer. These evidences lead us to believe that **the users do not know how to create a red color based on other primary colors**.

Since we have produced the mapping between given values and color bins, we can perform a qualitative analysis of each answer pair's colors names while comparing it to the expected ones. The expected pair was Magenta-Yellow: among 62 answers from the online users to question 8, **we conclude that the colors Magenta and Yellow were among the least indicated ones**, being Magenta indicated four times and five times. The most repeated color was Red, being indicated more than 70 times: there were 31 answer-pairs (50%) which contained the color Red twice! When given by the user, the Magenta was combined with Green and Red, while Yellow was combined with Red, Green and Blue; however, the specific pair Magenta-Yellow was never indicated.

2) Cyan + Yellow = Green: Both the laboratory and online users provided their answers on similar pairs, comprised between **blue shades and yellow ones**; the most repeated color pairs were, in fact, Blue shades combined with tones of green and yellow, as observable on Figure 8. Additionally, only two laboratory users (from twenty-three which indicated a pair of two answers - 8.70% of the users) and six online users (from seventy-one - 8.45% of the users) indicated an answer pair which did not even nearly approximate to the desired hues.

The expected pair was Cyan and Yellow: among 71 answers from the online users to question 17, **we conclude that the color Yellow was among the least indicated ones**; however, users have diverged a little when indicating the cyan color, since **the users have answered with colors varying between Blue, Navy-Blue, Sky-Blue, Light-Blue and Teal**, with frequencies of 37, 11, 3, 1 and 1 respectively, while Cyan was only referred 4 times. Though, users have referenced a lot more times the Green color, which is the presented one: 77 times to be more precise, with **53 answer-pairs formed with a Blue shade and Green** (75% of the totality of pairs) and 10 pairs which contained the Green color twice. If analyzed, the CIE Chromaticity Diagram in Figure 8 consolidates this idea, revealing the dispersion of blue derived colors' usage, but a slight concentration of answers in the neighborhood of Green and Yellow. Regarding the latter, it was combined 4 times with blue shades which is another indicator that **users presented strong Blue-Green/Yellow blending capabilities**.

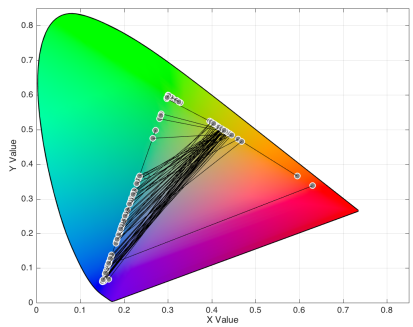

Fig. 8. Online: Question 17, Regular users.

E. Color Blending Effort

We have collected from the users the easiness of blending two shades of colors, when concluding each answer. This easiness is reflected in an numerical scale, ranged from one to five, being one the hardest level of effort, and five the easiest level of effort. The results are in agreement with the ones collected and concluded before: the **blending of Red-Green resulting in Yellow**, **blending of Red-Yellow and Green-Magenta, both resulting in Orange** and the **blending of Cyan-Yellow resulting in Green** were the mixtures which, according to the users' answers, had the lowest effort of blending, consequently the highest rating of easiness; particularly for the mixture of Red and Yellow, resulting in Orange, it provides the best results in both question types. Contrary, the questions which had the lowest results were almost always the ones in which the user was inquired about the result of given blending basis.

It is interesting to observe that the results for the generality of questions encircle the rating of 2 and 3 in the interval $[1;5]$, which could indicate that **apart from the obvious color blendings which proved to yield the best results, all the other blendings show nor an easy or hard level of effort to mix the basis**. These results are confirmed by the online responses, by running a Friedman Test which indicates us that there are no statistically significant differences ($p < 0.05$) when comparing the results from the laboratory environment, with the online results.

F. Demographic Results

We performed a set of statistic tests to realize if there is some kind of connection between the demographic information of our users and the answers given by them, for example a relation between the answers given by the male and female users, or even if the younger users interpret the color blendings differently from the older ones.

We have separated the users among some demographic groups: regarding the age division we have created six groups (**users aged below 20 years, aged between 20-29, 30-39, 40-49, 50-59, and above 60 years**), and for the gender we have divided the dataset according to **female, male and other gendered users**. These groups were, then, iterated by question, performing a *Spearman's* Rank-Order Correlation test between the ages and genders, with the answers. To test the differences between age groups, we have executed a *Wilcoxon* Signed-rank Test ($p < 0.05$) in order to detect significant differences which exist, between results for each color model, by age group.

The test reveals that the age group which was always present in each difference, for the selected questions, was **the users aged between 40 and 49 years**. However, **there is insufficient information for us to confirm that this age group presents much different values from the others**, since these questions present a minority of the sample. It is also **not possible to affirm that there are significant differences between all age groups**, because there are more questions which do not present any difference *per* color model and the analyzed differences include only a subset of all age groups. Hence, **we did not found relevant differences between age groups**.

Equally to age groups' analysis, we have executed a Wilcoxon Signed-rank Test ($p < 0.05$) in order to detect significant differences which exist, between results for each color model, by gender group. The test reveals that there is a mild difference between the results from the Female gendered users, and the Male ones. It is not an absolute difference, since **not every question had presented significant differences between gender groups**; it is **not possible to formulate a conclusion about which gender generates the wider or closer distances to the ideal answer (according to each Model)**, since there is no visible pattern about this subject. There were no significant differences between female or male users, and the *other* gendered users: this could due to the lack of a significant user sample relating the later gender.

G. Guidelines for Color Blending Usage

The following list enumerates a set of guidelines which we are able to formulate, based on our user study, about the usage of color blending when sketching visualizations:

1) To maximize the compatibility with the user's expectations, the **chosen** color blends must present Orange shades based on mixing Red and Yellow.
2) Color blends which present a resulting color of Green are easier for the user.
3) The Yellow color provides very good results, when blending Red and Yellow.
4) The CMYK Color Model is the one which has the best results, according to user's expectations.
5) To maximize the compatibility with the user's expectations, **avoid** presenting colors which are near Blue shades.
6) When presenting Red shades, the best blending-basis which generates it is Red and Magenta.
7) Presenting the resultant color from a given blending, generally, produces better results than asking the user to blend the color according to his mental model of color.
8) The color model which best resembles the users' expectation is a subtractive one.
9) If the usage of CMYK Color Model is not desired, the usage of HSV, RGB and CIE-L*a*b* can produce similar results between each other.

10) Avoid using the CIE-L*C*h* Color Model when blending colors according to the users' expectations.

V. Conclusion

We aimed to answer three questions: **Q1:** Which Color Model meets best the users' expectations, when blending two colors?, **Q2:** Is there evidence of a spatial arrangement of colors, when users are indicating possible color mixtures' results?, and **Q3:** Are there evidences from differences across demographic groups, such as the age or gender?

We gathered a considerable amount of 259 users, which helped us study which color model yielded the best results. We have found that CMYK is model that best resembles the users' expectations, while the orange, green and yellow color-blending results are the ones which generate shorter distances to between the expected colors and the ones indicated by the users. On the other hand, the CIE-L*C*h* Color Model is the one which is farther apart from the users' mental model of color.

It was interesting to observe that there is a mild indicator that it exists a difference between the responses from Female and Male users. However, since our user sample was not complete enough to determine this difference, and not every color blending has presented statistically significant differences between genders, there is no substantial ground-truth to formulate any a formal research conclusion.

We have also formulated a set of guidelines which could be followed when using color blending to convey information, which gathered the conclusions from the all the study results' analysis and summarized it in rule-of-thumbs to follow.

With this Master Thesis, we have set some implications for the Information Visualization field of research, from the Color Model to use when presenting color blendings, to what should be asked the user (either the blending-basis or the result of the blending) in order to maximize the success rate of each visualization, following the color blendings which yield the best results at the same time they are consistent with the human color perception.

We consider that there were some limitations to this Master Thesis, regarding the lack of users keen to participate in our study, since we believe that the greater the user sample, the better the results of these user study would have been. Another constraint was the location available to conduct the user laboratory study, since the study had to be realized in rotating locations because there was no constant, quiet and well-located space for us to fulfill the user sessions. The absence of this factor would cause the broadcasting of the laboratory sessions to be much easier than what, in fact, was. Lastly, we believe that this research could have benefited if it had had another disclosure to users: if we had had the opportunity to broadcast this user study in other countries, we could have attained a sample of users culturally quite different, which could have enhanced the analysis of other cultural differences.

A. Future Work

This field of research has proved itself to have a tremendous potential, whereby there is a large set of questions which remain unanswered. Since the size of the user sample is a major concern when conducting user study like the one conducted by us, and calibration is an unavoidable issue, it would be interesting to conceive a remote calibration system which would be capable of rendering the web page container according to the user's LCD Display calibration.

Comparing the results given with commonly-named colors was an important part of our analysis. Nonetheless, the comparison against the XKCD's Color Bins was not seamless: the generated bins and areas of coverage of each named color were not perfect, so it would be an interesting research topic to provide a comprehensive study about the naming of colors, in a laboratory environment.

Since it was mildly observable in our user study the theory that there is a difference in results between gender groups, further investigation could deepen if there is, in fact, any plausible difference between genders or age groups.

Finally, respecting the color blendings, it could be further analyzed and deepen the relationship of human color perception with the Blue color: although it was the one which produced the weakest results of this color study, it still exerts some kind of influence on mental models of color of our users.

References

[1] M. Chirimuuta, "The Metaphysical Significance of Colour Categorization," in *Colour Studies: a Broad Spectrum*. Anderson, Biggam, Hough & Kay (eds.), 2014, pp. 1–29.

[2] S. Gama and D. Gonçalves, "Studying Color Blending Perception for Data Visualization," in *EuroVis - Short Papers*, N. Elmqvist, M. Hlawitschka, and J. Kennedy, Eds. The Eurographics Association, 2014.

[3] S. Gama and D. Gonçalves, "Studying Color Blending for Visualizing Social Artifacts," in *Encontro Português de Computação Gráfica, 2014. EPCG2014 - 21*, 2014.

[4] S. Gama and D. Gonçalves, "Studying the perception of color components' relative amounts in blended colors," in *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, ser. NordiCHI '14. New York, NY, USA: ACM, 2014, pp. 1015–1018. [Online]. Available: http://doi.acm.org/10.1145/2639189.2670264

[5] N. Gossett and B. Chen, "Paint inspired color mixing and compositing for visualization," in *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, 2004, pp. 113–118.

[6] P. Ginter, J. Jones, and S. Hoda, "True colors," *International journal of surgical pathology*, vol. 19, no. 4, pp. 494–496, 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23550367

[7] C. Ware, *Information Visualisation: Perception for Design*, 2nd ed., S. Card, J. Grudin, and J. Nielsen, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2012.

[8] L. Zhou and C. D. Hansen, "A survey of colormaps in visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 8, pp. 2051–2069, Aug 2016.

[9] F. Stoffel, H. Janetzko, and F. Mansmann, "Proportions in categorical and geographic data: Visualizing the results of political elections," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ser. AVI '12. New York, NY, USA: ACM, 2012, pp. 457–464. [Online]. Available: http://doi.acm.org/10.1145/2254556.2254644

[10] D. de Alwis and C. Kon, "A new way to use the ishihara test," *Journal of Neurology*, vol. 239, no. 8, pp. 451–454, 1992. [Online]. Available: http://dx.doi.org/10.1007/BF00856810