https://doi.org/10.2352/CIC.2022.30.1.42

Representative Color of Skin Tones and Natural Objects

Dara Dimoff; Rochester Institute of Technology; Rochester, NY
Susan Farnand; Rochester Institute of Technology; Rochester, NY

Abstract

Color enables humans to readily extract features of an object, leading us to describe tomatoes and apples as "red" despite the presence of other colors. How observers accomplish this is not well understood. In this study, we present observers with rapidly presented stimuli at varying levels of context. Observers were asked to select the color that best represents the image from eight options. We found that observers tended to select progressively lighter or darker colors as more context was introduced, although whether the representative color choice became darker or lighter varied from image to image. This is likely a result of observers discounting achromatic cues (i.e.: specular highlights, shadows) as context is revealed, but why images were treated inconsistently requires further investigation. Observer responses were noisily distributed. These results shed light on how observers characterize the color of multicolored objects.

Introduction

Color enables us to make valuable judgments about the features of an object. Especially in the natural world, objects are rarely a solid color, but are composed of multiple colors as well as shadows and specular highlights. Figure 1 shows an image of a tomato whose color tells us about its ripeness or what it might taste like.

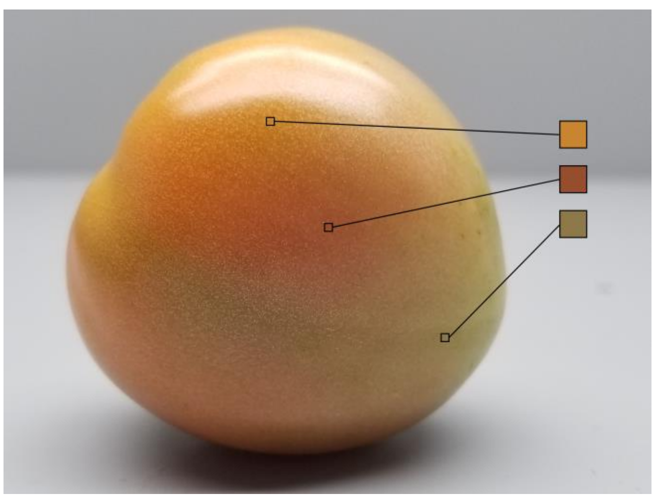

Figure 1. Complex objects such as tomatoes can have a variety of colors.

Representative color can be defined as the color that best characterizes an object. How do humans determine what the representative color of a multicolored object is, such as the tomato in Figure 1? An observer might focus on the principal color of the object and omit accenting colors as well as shadows and specular highlights to make a judgment. Alternatively, one might attempt to "average" the colors present the same way a computer could calculate the average RGB value in an image. While this question is relatively unexplored, a recent study by Virtanen et al [1] showed that observers tend to perform averaging when asked if a simple stimulus was "bluer" or "yellower." In this study, we further explore this question to gain understanding for what factors are considered when making a judgment for the representative color of an object.

Methodology

This experiment was designed as an extension of Carpenter's [2] study that focused on the representative color of tomatoes for use in the agricultural industry. We expand on this prior study, instead with a focus on skin tones of human faces. Additionally, to address the limited observer availability during the COVID-19 pandemic, two identical experiments were run and their results compared: one online and one in a controlled onsite environment.

In these two experiments, observers were presented with an image of an object for one second. Then, out of a selection of eight color patches, they were asked to select the one that they considered to be the best representative color of the object.

Observers

A total of 40 observers participated online and 16 observers participated in the onsite controlled environment. Some observers that participated in the onsite experiment also participated in the online experiment, in which case, it was noted which experiment the observer took first. Observers taking the experiment online received additional instructions to disable screen filters (ie. blue light filters) and report their viewing conditions (ie. indoor dim). Basic background data was collected, including age, gender identity, whether the observer has experience in imaging, and whether the observer has a color vision deficiency. The experiment generally took 10-15 minutes to complete.

An important limitation of this study to note is that observer ethnicities were not collected. Observers were recruited through a university recruitment system which did not include ethnic data of subjects. This oversight will be corrected in future data collection.

Apparatus

The onsite experiment was conducted in a dark room on an Eizo CG248. The screen size was 23.8" and the screen resolution was 3840x2160 with a refresh rate of 60 Hz. The monitor's settings were calibrated to sRGB with a D65 measured white point of XYZ = (99.71, 100.3, 116.7). The background grey was selected to be RGB = (171, 171, 171) with a measured value of XYZ = (41.05, 41.29, 48.12). The experiment was accessed via a webpage built using PsychoPY. Observers sat 70 cm away from the monitor.

Stimuli

Stimuli were generated using MATLAB (version 2022Ra). 20 photos were used to produce stimuli for this experiment (12 faces, 3 wines, 2 tomatoes, 1 pepper, 1 sky, 1 grass). The photos of faces contained subjects of the following ethnicities: 3 African Americans, 3 Asians, 3 Indians, and 3 Caucasians. To understand what strategies humans use to determine the representative color of an object, we must first understand whether context or smaller features play a role. For example, if the observer does not know that dark pixels are shadows or that pink pixels are lips, then their choice for the best representative color of an image may change. To test this, four versions of each photo were generated: a full resolution image (768x768 pixels), a downsampled 64x64 version, a downsampled 16x16 version, and a randomized shuffle of the 16x16 image (Figure 2). Downsampled images were not generated using nearest-neighbor averaging because this would create pixel colors not seen in the original image; instead, every nth pixel of the full resolution image was retrieved to construct the downsampled image, where n = 48 in the 16x16 and n = 12 in the 64x64.

Figure 2. Four stimuli were generated from a 768x768 photo. From left to right: shuffled, downsampled 16x16, downsampled 64x64, original photo.

Eight color patches were generated for the observer to select as the color that best represents the image they are shown. This experiment utilizes the observer's color memory by showing the stimulus first and then the color patches to choose from. In an earlier study, Bartleson [3] found that observers tend to remember colors with more saturation and lightness than the original stimulus. This suggests that observers might be discounting achromatic features, such as shadows and specular highlights. This was taken into account when generating the eight color patches, as shown in the process below. Color patches were obtained from the downsampled 16x16 image because the pixels present in that image would be present in all other versions of the photo. The process for generating color choices was as follows:

1. Select the most saturated pixel, the most chromatic pixel, and calculate the image's average color. Add all three to the set of color choices.
2. Segment the image into regions (for human faces, the regions were: left cheek, right cheek, chin, neck, and nose). Repeat step 1 for each region.
3. Remove colors that are closer than 3 ΔE00 in CIELAB space to any other color in the set. This threshold was selected because a difference of about 2.5 – 3.0 ΔE00 has been determined to be just perceptible in images.
4. If more than 8 colors remain, alternate between the darkest and lightest colors (begin with whichever is furthest from the average) and iteratively delete colors until 8 remain.
5. If there are fewer than 8 colors in the set, iteratively add colors that are midpoints between the colors in the set. Repeat steps 3 and 4 on the new colors until 8 remain.

For this process, pixels that were not part of the object of interest (ie. the background) were not included when averaging colors or selecting most chromatic/saturated colors. Figure 3 shows the eight colors that were generated for the stimuli in Figure 2.

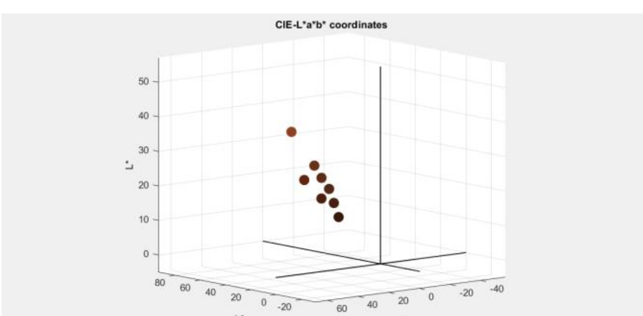

Figure 3. CIELAB coordinates of eight color choices.

Procedure

The experiment was divided into four sections: shuffled stimuli, 16x16 stimuli, 64x64 stimuli, and full resolution images. The four sections were always presented in the same order so that observers would be unaware of the context of the objects in the shuffled and lower resolution images. Each section had 20 images for a total of 80 trials. Observers were presented with 4 training images before having their responses recorded. Images within each section were presented in a random order. Each trial began with a blank screen for 1 second, followed by the stimulus for 1 second, then the blank screen again for 1 second. Then, the observer was shown eight color patches, displayed in a random order each time, from which they selected one using a mouse cursor. The timing of each trial is shown in Figure 4. Maule et al [4] have shown that this process of rapid presentation is sufficient for observers to reproduce average hue over a given array.

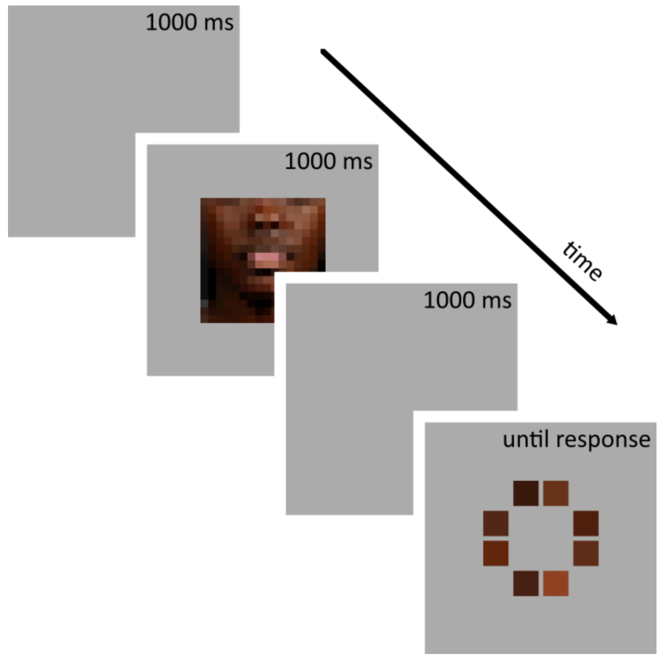

Figure 4. Example timing of a single trial.

Results

Figure 5 shows the observers' mean color choice for one set of photos in the online experiment. The solid color bordering each image is the average selection observers made for the corresponding image. Notice that the mean color choice becomes progressively darker as more context is added. This could be attributed to the amount of skin that contains specular highlights; in the scrambled image, that context is unknown, and so a lighter color might be selected in that image. This trend, however, was inconsistent across different photographs. See Figure 6 for the opposite effect; observers' choices become progressively lighter as more context was introduced. One might attribute this difference to the amount of skin that is cast in shadow, and that once that context is understood, the observer removes the shadow from their judgments. Most photos of subjects other than Caucasians exhibited this effect: the image became either progressively lighter or darker.

Observer responses to photos of Caucasian subjects exhibited less variation in lightness and more variation in chroma as more context was introduced. Figure 7 shows the average results of a

Caucasian subject with significant pinkish blotches. Observer responses decreased in chroma as context was introduced, suggesting the possibility that observers place less value on temporary skin colorings when judging representative color, such as the pinkish tints from blush or sunburn. It is also worth noting that color options contained less variation in lightness for Caucasian photos compared to photos of other subjects. $L*$ values for color choices for Figure 5, Figure 6, and Figure 7 ranged from 13.4 to 37.6, 31.7 to 57.4, and 52.4 to 66.4, respectively.

Figure 8, Figure 9, and Figure 10 show nested histograms of observer responses per color patch. Blue bars represent total responses for the corresponding color across all trials, while the colored bars nested within show the total number of times that color was selected based on the trial block (from left to right: shuffled, 16x16 downsampled, 64x64 downsampled, full resolution). Notice that observer responses were distributed noisily across all images. Observing overall average choices does not seem to reveal much, however, observing averages based on context reveals some trends. In Figure 9, for example, color #1 and color #8 see a massive change in responses when looking at the scrambled image versus the images that include the context of a human face.

Figure 11 and Figure 12 give a similar overview of one of the natural objects (tomato) included in the study. In this study, full resolution images of natural objects show the entire object rather than a small region, which must be considered when analyzing the results. In this set of images, it is probably not obvious to the observer that they're looking at a tomato until the full context image is shown. Observer responses progressively increased in lightness as context was shown by approximately 1 $L*$ per level of context. Observer responses varied in $a*$, with trial 3 having the "reddest" response and trial 4 having the "greenest" response. We can speculate that the observer may be discounting the red accents in the full resolution image.

It is also important to consider the limitations of presenting only eight colors to the observer. Eight choices were presented because more than eight might overwhelm the observer, however, with only eight choices, observer selections are limited in lightness and chroma. An observer that might want a redder color choice for the tomato will be limited in lightness choices among the color choices that have a higher $a*$ value.

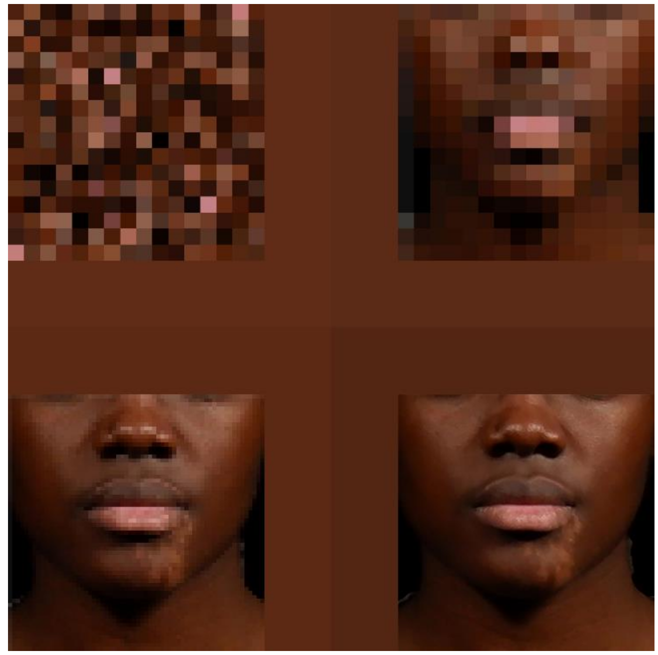

Figure 5. Average color response for one set of images of an African American subject. The strip of color surrounding each image indicates the average response for that image. Observer responses became progressively darker as more context was introduced.

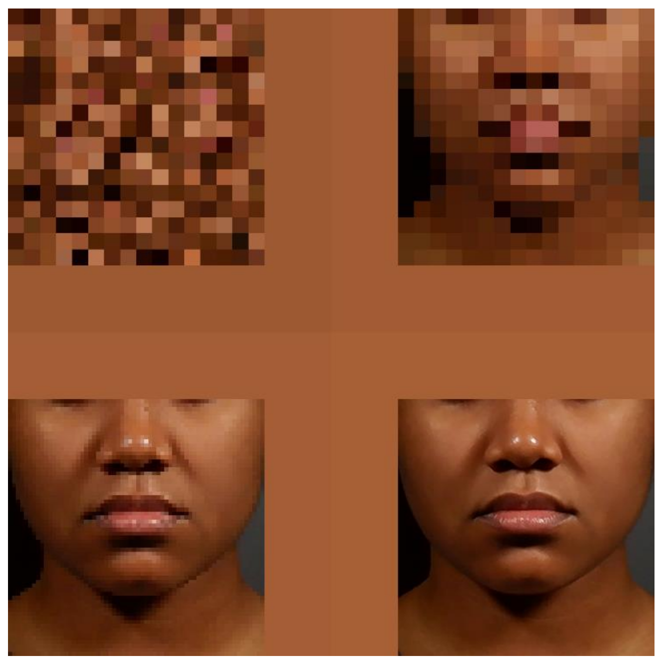

Figure 6. Average color response for one set of images of an Indian subject. Observer responses become progressively lighter as more context was introduced.

Figure 7. Average color response for one set of images of a Caucasian subject. Observer responses were highly chromatic for the shuffled image, but once the context of a human face was introduced, responses became less chromatic.

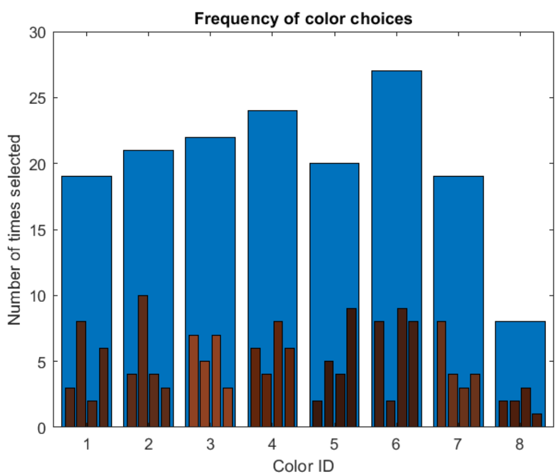

Figure 8. Histogram of observer responses corresponding to Figure 5. The colors of the bars contained within the blue bars are equivalent to the RGB values shown to the observer.

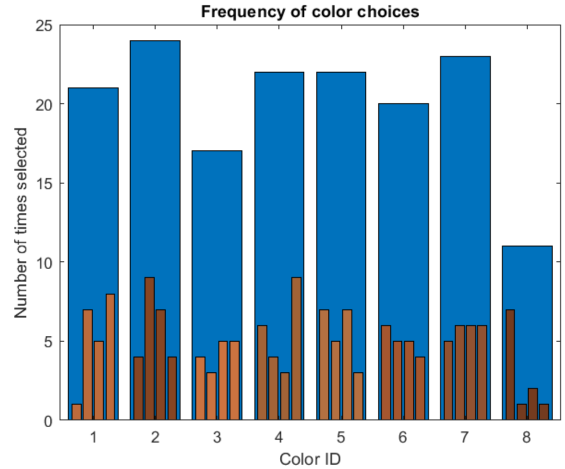

Figure 9. Histogram of observer responses corresponding to Figure 6. Notice that color #1 was rarely selected for the shuffled image, but selected frequently once the context of the human face was introduced. The opposite phenomenon occurs for color #8, the darkest choice in the set.

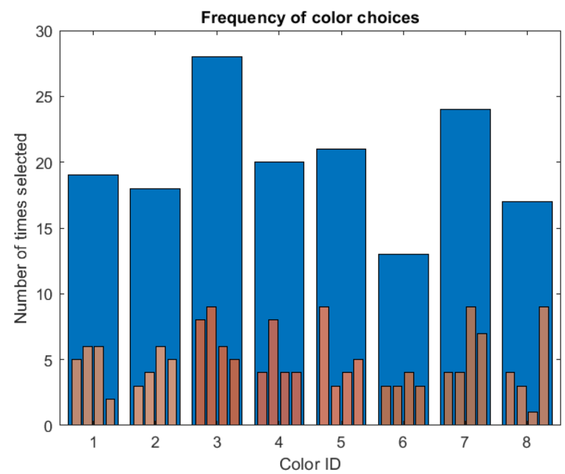

Figure 10. Histogram of observer responses corresponding to Figure 7. Color #5, the most chromatic color in the set, is predominantly selected for the shuffled image.

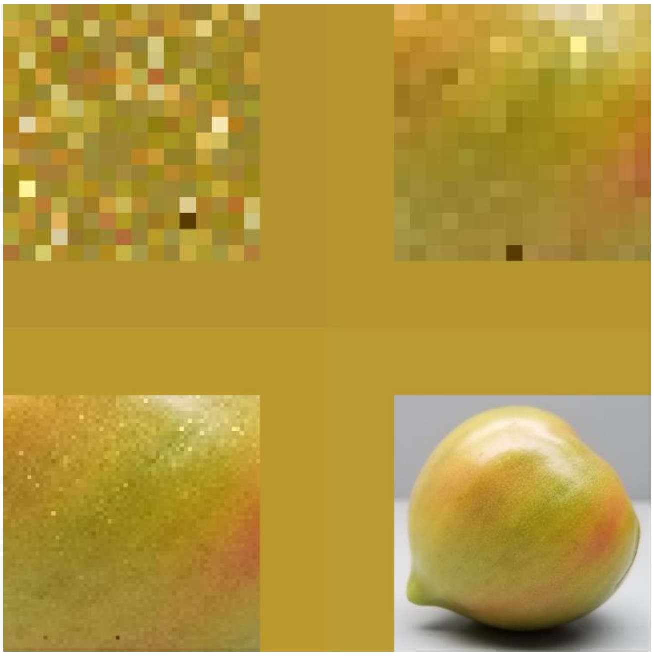

Figure 11. Average color responses for one set of images of a tomato. Observer responses became progressively lighter, but varied on the a dimension.*

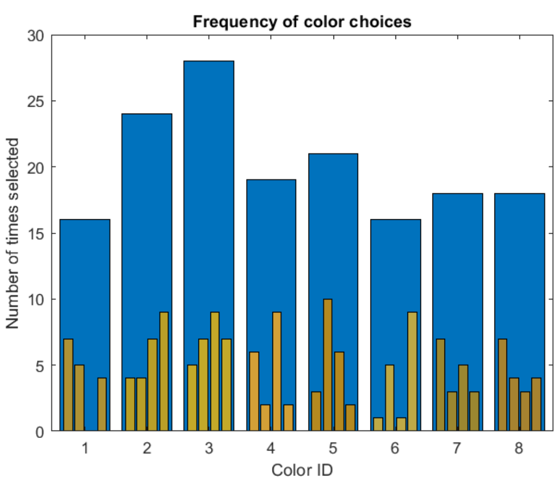

Figure 12. Histogram of observer responses corresponding to Figure 11. Notice the large increase in popularity for color #4 in the downsampled 64x64 image and the increase in color #6 for the full resolution image.

It is worth noting that observers selected color patches in the same position as their previous selection more often than color patches in other positions. With eight color patches, one would expect that colors in the same position as the previous trial would be selected 12.5% of the time, however, observers selected the color from the same position as their previous choice 14.5% of the time. While this increase was modest, it suggests that observers might continue looking at the portion of the screen of their last selected color, causing them to notice the color patch under their mouse cursor sooner than the rest, or they may feel less inclined to move their mouse if the color patch under the mouse seemed reasonable.

In the onsite experiment (n = 16), observers selected the most chromatic color choice 15.42% of the time and the most saturated color choice 10.33% of the time. Observers in the online experiment (n = 40) selected the most chromatic color choice 12.78% of the time and the most saturated color choice 12.81% of the time. This is contrary to our prediction that observers will be drawn to the most saturated/chromatic pixel over the rest. The difference between the two groups could be due to the small sample size for the onsite experiment.

After taking the experiment, observers reported that they felt one second was insufficient to observe the image to accurately select a representative color. While the experiment performed by Maule et al [4] showed that half a second should be sufficient, it could be argued that the more complex nature of the images in this experiment would require additional time. A follow-up experiment will assess the impact of increased observation time on color selections. The follow-up experiment will also assess observer repeatability of their color choices.

Conclusions

While most images followed a trend of being viewed as either progressively lighter or progressively darker as more context was provided, it is unclear what factors determine why some images increase in representative lightness as context is introduced and others exhibit the opposite effect. It is possible that photos which exhibited higher proportions of perceived shadow versus perceived specular highlights is responsible for the increase in lightness from observer judgments, however, this requires further investigation.

It is important to note that observer responses were distributed noisily. This is consistent with the findings of Maule et al [4]. The reason for such high disagreement between observers is unclear but may be related to individual differences in strategies for determining representative color.

Acknowledgements

This material is based upon work supported by the AdvanceRIT program, originally funded through the National Science Foundation under Award No. HRD-1209115.

References

[1] L.S. Virtanen, M. Olkkonen, T.P. Saarela, "Color ensembles: Sampling and averaging special hue distributions," Journal of Vision, 20, 5 (2020).

[2] Carpenter, K.M. (2021). Assessing the Use of Smartphones in Agriculture (Doctoral dissertation, Rochester Institute of Technology).

[3] C.J. Bartleson, "Memory Colors of Familiar Objects*," Journal of the Optical Society of America, 50, 1 (1960).

[4] J. Maule, A. Franklin, "Accurate rapid averaging of multihued ensembles is due to limited capacity subsampling mechanism," Journal of the Optical Society of America, 33, 3 (2016).