Color names affect the precision of memorized hues. The effect of increased color name distinctiveness

Zuzanna Skóra and Jon Y. Hardeberg

Colourlab, Department of Computer Science, NTNU - Norwegian University of Science and Technology, Gjøvik, Norway

Abstract

Color names influence the memory of a specific hue by shifting it closer to the prototypical color associated with the name. Typically, in studies regarding color naming and memory maximally seven traditional color names are used. We tested whether the distinctiveness of the color name can modify the shift of the memorized hue by introducing unique color names for all hues presented in the experiment. Specifically, we gave unique names to boundary colors, defined, based on previous experiments, as hues that are equally likely named using two traditional names. Central colors are consistently named with one of the seven traditional color names. We observed that more distinctive names were associated with a poorer color recall performance than the traditional names. We propose it can be the result of long-term memory overload caused by having to memorize new color names. At the same time, when distinctive names were provided for boundary colors, the difference in the ability to accurately recall a hue between the boundary and central colors was smaller than when only traditional names were used. Moreover, when using only the traditional color names for boundary colors, we observed a consistent shift in the recalled hue towards the direction of the color name that accompanied it. This study should be considered a preliminary one. Follow-up experiments should be conducted to assess the exact cause of the observed differences. As the next step, we propose to test whether alternate labels to Pantone would lead to similar results.

Keywords

Color Memory, Color Name, Short-Term Memory.

1. Introduction

The relationship between language and color has been studied for a while now. Most studies focused on differences between languages in the number of hues they distinguish [1] and the impact of these differences on observed behavior while performing a color perception task [2]. However, studying differences in hue perception between different languages does not allow us to differentiate between the effects of more distinctive color names and the impact of culture and experience. Learning that it may be the color label distinctiveness, for example using two labels to distinguish between a blue that falls closer to green and one closer to pure blue, that improves color differentiation and not necessarily having grown up in a culture that made such a distinction, could be used to improve color discrimination and recall in the general population.

In this paper, we investigate the effects of newly-acquired labels on color memory over the short term. There is already some evidence suggesting that color names, in general, affect our perception and memory of colors [3, 4, 5, 6]. These studies show that we perceive and memorize colors biased towards their category centers (prototypes) and this bias is stronger when we explicitly label each hue. One of the newest studies [5] showed that if we constrain the label distinctiveness - from 7 basic labels to 4 or 2 labels used for the same color spectrum - the recalled hues are clustered around those labels. It suggests that labels may to some extent control color memory and likely also perception [3].

The 11th Colour and Visual Computing Symposium, September 08-09 2022, Gjøvik, Norway
EMAIL: zuzanna.skora@gmail.com (A. 1); jon.hardeberg@ntnu.no (A. 2)
ORCID: 0000-0001-6018-8031 (A. 1); 0000-0003-1150-2498 (A. 2)

Here, we followed the idea behind the study by Souza et al. [5] but instead of constraining the labels we provided participants with additional labels to distinguish between hues at the border of two basic color categories (central colors), henceforth referred to as boundary colors. These hues are typically the most difficult to remember [3, 5], probably because the basic labels, that participants usually operate with, refer to the central colors, pulling the responses towards them. Based on this observation, we tested if providing additional labels can lower the degree of bias towards the color category center.

In addition to providing new labels for boundary colors, we were also interested in the impact of labels associated with central colors on boundary colors. In previous studies in which participants were naming the hues themselves [5] boundary colors were equally likely to be given a label from the neighboring central color on the left or right, as positioned on the color spectrum. In our experiment, for each boundary color presented, participants could hear either the label associated with the central color on the left or right. We aimed to assess the impact of labels on color memory by testing whether the category bias can pull boundary colors in the direction of the label.

The experimental setup and chosen hues were kept as similar to the study by Souza et al. [5] as possible. One reason for this decision was that we not only introduced novel conditions (more distinctive labels and boundary color label manipulation) but also, instead of asking participants to name the hues themselves, we played pre-recorded labels while presenting each hue. As a standard in research on color memory, the impact of the color label is studied by asking participants to provide color names. However, there are already some studies showing the effects of hearing color names on color perception [7].

Another novel factor in this study is a more precise measurement of color recall than typically used in color naming studies. Typically, in studies on the effects of language on color perception the task and the outcome measure capture the notion of color performance quite broadly. The task is typically a recognition task in which the speed of color recognition is the outcome measure [6, 2]. In the following experiment, we use a task recently developed in the area of working memory research. It is a continuous recall task and its outcome is a recall error. In the case of color, recall error provides information on the circular distance between the presented (target) color and the color recalled by the participant. This measure is already more direct than response time and more information can be acquired based on it by implementing some of the models proposed to explain the components of memory performance (e.g. mixture model which assumes that recall error is a combination of the probability of remembering a color, the precision of the remembered color and guessing rate [8]). The more precise measure may help to find differences that could have been overlooked with the indirect measures applied so far [9].

2. Methodology

2.1. Participants

We tested 13 participants (2 females), with a mean age of 35.4 ranging between 24 and 59, and no known color deficiency. Each participant provided informed consent to participate in the study before they began and was not reimbursed.

2.2. Equipment

For the experimental task, we adapted the Matlab code provided by Souza et al. [5]. Link to the original code: https://osf.io/mqg4k/. The experiment was displayed on a 24.1'' EIZO CG241W monitor with the viewing distance unconstrained. Participants were tested individually with a constant light level in the room. The background throughout the whole experiment was a uniform grey (RGB: 128 128 128). The selected colors were sampled from 360 values evenly distributed along a circle in the CIELAB color space with $L^* = 70$, $a^* = 20$, $b^* = 38$, assuming sRGB.

2.3. Procedure

At the beginning of the experiment, participants read through instructions. As the color stimuli, we used values extracted from the study by Souza et al. [5]. First, we extracted seven RGB values that were associated with the highest agreement about the label and named them category centers. Next, we extracted seven more RGB values associated with the lowest agreement about the label (e.g., a value between orange and yellow that was just as likely labeled orange or yellow) which served as our boundary colors. Figure 1 presents the proportion of providing a specific label for a given color on the color spectrum which was the basis of our stimuli selection.

The experimental task consisted of two conditions that differed only in the labels played while viewing each to-be-remembered color. In the traditional condition, participants heard seven basic labels. Central colors were always associated with the same label; however, boundary colors could be presented with either the label from the central color to the left or the right. For the second condition, which consisted of distinctive labels for both central and boundary colors we used the Pantone color names based on the FHI (FASHION, HOME + INTERIOR) System (https://connect.pantone.com/). Based on these values we found corresponding Pantone color names. Table 1 contains the information on each color presented in the experiment.

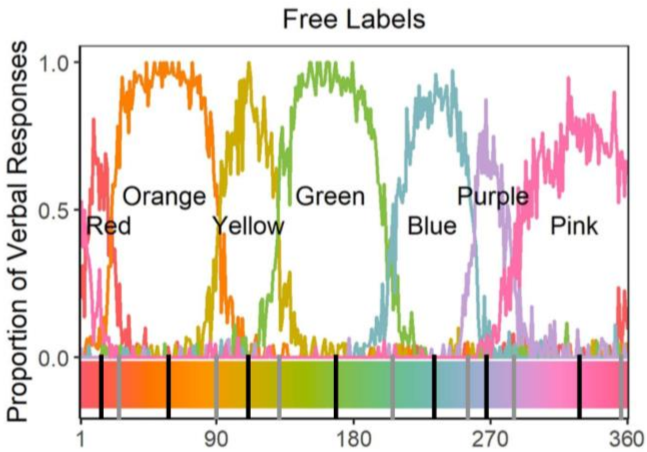

Figure 1: Verbal labels recorded during the memorizing phase of Experiment 1 in the paper by Souza et al. [5]. Each line represents the number of times a given label was provided for a presented color. Note that in the study from which this data comes, each color on the spectrum was equally likely to be presented, unlike in the current study for which we only chose 14 colors. Reprinted from Souza et al. [5] (CC-BY license) with a modification in the form of vertical lines indicating the hues chosen as central (black) and boundary (grey) colors for the current experiment.

Before moving to the main task, we displayed all 14 colors, one at a time, together with their unique Pantone label. This was done to allow participants to familiarize themselves with the new label-color association and to have a chance to not only hear the color label but also see how it is written. In the main task, each experimental trial consisted of two phases: study and test, see Fig. 2 for the flow of a trial. During the study phase, participants viewed four different colors presented sequentially. Two of them were central and two boundary colors, however, the sequence was shuffled randomly. The position of the first color was randomly sampled out of four possible starting points, and the three other colors followed in a clockwise fashion.

Next, the memory of each color was tested, again, choosing the starting color at random and then following in a clockwise fashion. At the beginning of the test phase, a color wheel was displayed that was randomly rotated from trial to trial. The task was to choose from the color wheel the color previously presented on the spot to which the arrow now pointed. By moving the mouse around the color wheel participants could see the exact color value and click on the spot that according to them corresponds best to the memorized value. The experiment consisted of two separate blocks, one per condition. We varied the order of the conditions between participants to be able to measure the possible influence of one condition on the other. Each participant completed two practice trials before each of the two blocks and 53 test trials per block.

Table 1
Angular values extracted from Souza et al. [5], their translation to RGB values, and their classification to different conditions introduced in the experiment.

angle	R	G	B	type	traditional	Pantone	hue
5	255	90	100	boundary	*pink-red*	Coral Paradise	
9	255	91	92	central	*red*	Hot Coral	
20	255	96	70	boundary	*red-orange*	Vermillion Orange	
56	255	126	0	central	*orange*	Bright Marigold	
90	238	156	0	boundary	*orange-yellow*	Gold Fusion	
111	206	171	0	central	*yellow*	Daylily	
130	178	180	0	boundary	*yellow-green*	Grenoble Green	
165	132	189	68	central	*green*	Jasmine Green	
205	107	189	149	boundary	*green-blue*	Jade Cream	
232	125	183	189	central	*blue*	Amazonite	
260	171	169	211	boundary	*blue-purple*	Baby Lavender	
272	195	161	213	central	*purple*	Lavendula	
285	220	150	210	boundary	*purple-pink*	Lilac Chiffon	
325	255	112	169	central	*pink*	Pink Lemonade	

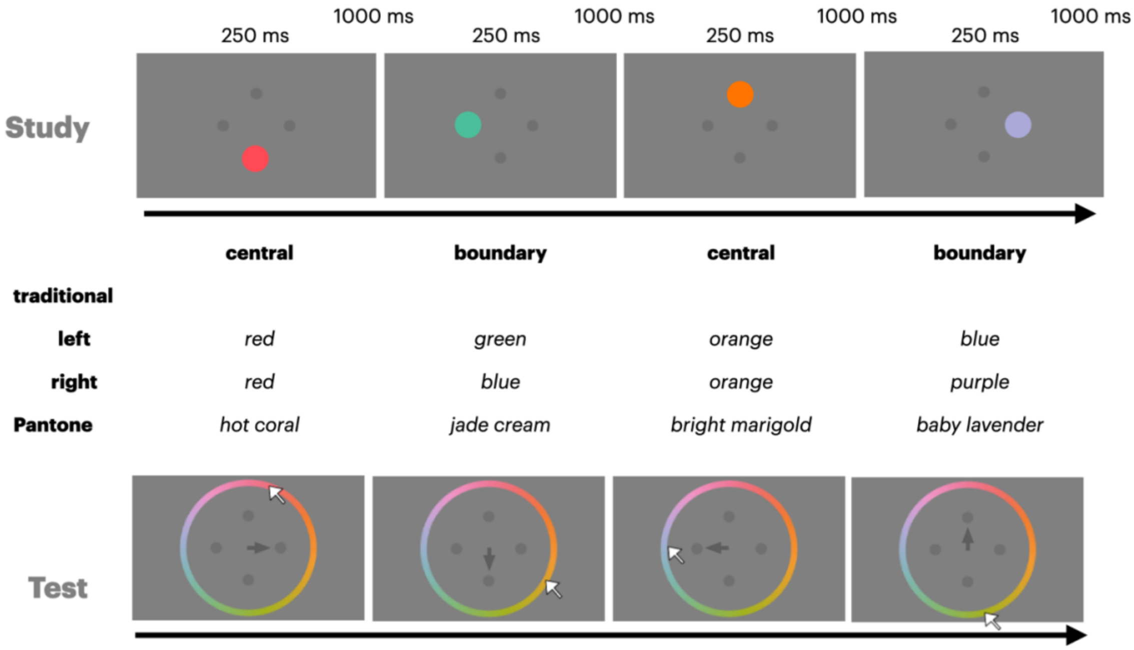

Figure 2: The flow of a trial during the experiment. Each trial began with a study phase in which each of the four colors was introduced. The position of the first color was randomized, and the other ones followed in a clock-wise manner. The presentation of each color during the study phase was accompanied by an audio recording of its label. The labels varied between traditional and Pantone conditions but also for boundary colors within the traditional condition. During the test phase, an arrow pointed to the previous position of each color in a random order.

2.4. Code availability

The adapted Matlab scripts to run the experimental task and analysis scripts for R can be found here: https://osf.io/cn95t/?view_only=7001b8e5d2d048f694c1335188e5d492

3. Results

The data analysis described below was performed in R [10]. To acquire mean estimates and the 95% credible intervals we used Bayesian multilevel models as applied in the *brms* package [11]. Similarly, to confidence intervals, credible intervals also summarize uncertainty related to the unknown parameters that we are trying to estimate, however, credible intervals are based on probability distribution. The credible interval is the range of values within the estimated posterior distribution created through the process of Bayesian inference. In short, a 95% credible interval is the central portion of the posterior distribution that contains 95% of the values.

3.1. Data overview – response distribution

When considering the response distribution, one important observation is that there seem to be differences between colors in how often they were recalled showing that we do observe color-specific bias. This is reflected in the shape of the distribution of each hue on the response spectrum. The flatter the distribution for a particular hue, the worse its memory. Note, that unlike in the previous experiments investigating the effects of color naming on color memory we did not present stimuli across the whole color spectrum (0 – 360) with equal probability. Instead, we chose 14 specific hues based on the verbal labeling data from experiment 1 in the paper by Souza et al. [5]. Additionally, the position of the hue-specific peak of the distribution in relation to the exact hue that was presented can inform us about the strength of the label bias between central and boundary colors (on Fig. 3 represented as black and grey vertical lines, respectively).

By just visually examining the difference in the response distribution, there does not seem to be any substantial difference between the conditions. To investigate this more thoroughly, we need to look more closely at the performance estimated using the recall task.

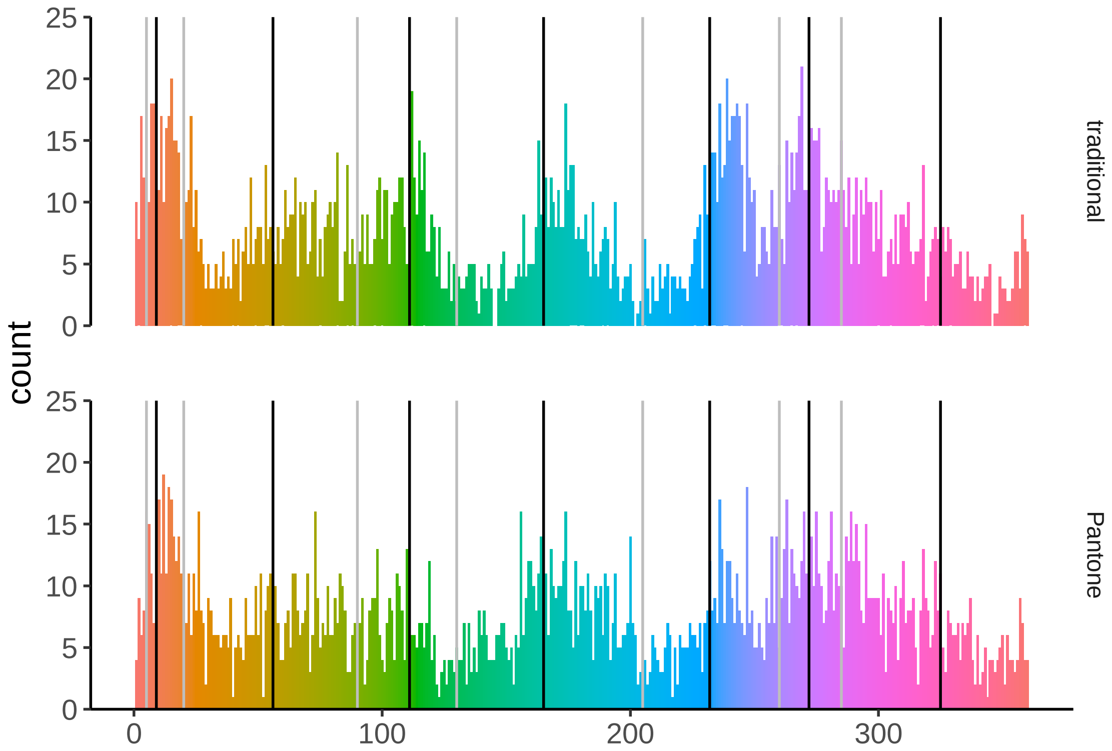

Figure 3: Distributions of responses in the traditional (upper) and Pantone (lower) conditions. The actual 14 presented colors are marked with lines, black for central colors and grey for boundary colors.

3.2. Estimating the difference between conditions with recall error

First, we observed that, overall, the Pantone condition was associated with higher recall error than the traditional condition (see Fig. 4). Moreover, in the traditional condition, boundary colors seem more difficult to recall than central colors. In the Pantone condition, the difference between central and boundary colors was more due to the wider credible interval for boundary condition as compared with central rather than mean estimates.

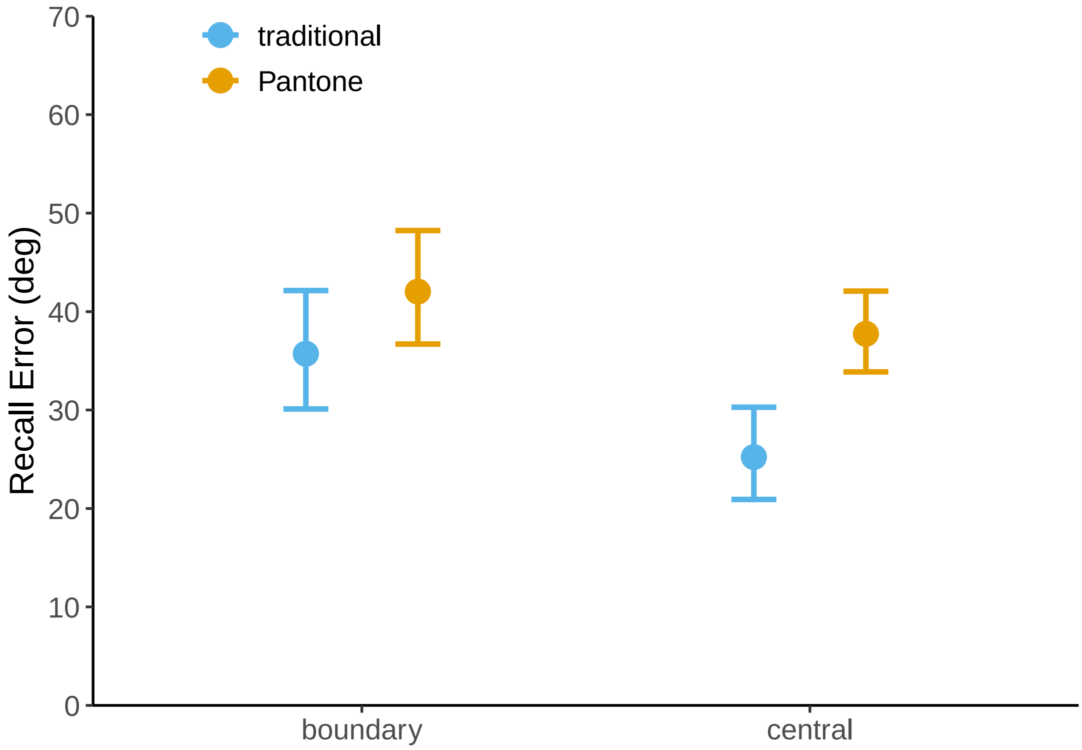

Figure 4: Estimated recall error for the labeling conditions and the color type conditions plotted together with 95% Credible Intervals (CIs) and individual participant means.

Pantone labels were overall more difficult probably because participants did not know them beforehand, as they likely did traditional labels, and some of them consisted of two words. Both of these factors make it more difficult to use the label as a long-term anchor for color memory. At the same time, the limited involvement of long-term memory may have resulted in less category bias. Pantone labels may have pulled the remembered hues less towards the color category center, and as a result, boundary and central colors shared more similar recall errors than in traditional condition.

3.3. Exploratory analysis of stimulus position effect

For more evidence in favor of differences in long-term memory involvement between traditional and Pantone conditions, we looked at the primacy and recency effects (see Fig. 5). These effects are typically observed during a sequential stimulus presentation. The stimulus presented as the first to memorize in the sequence is transferred to long-term memory and thus, we observe that it is recalled better than the following stimuli. This is the primacy effect. The recency effect is observed for the stimulus presented as the last in the sequence. It is still within the focus of attention and no other stimulus comes afterward therefore it is not prone to be overwritten. The recall for the last stimulus is typically higher than for the ones preceding it, however, this effect does not rely on long-term memory. If it is the long-term memory involvement that differentiates between traditional and Pantone labels in our experiment, we would expect to observe overall worse recall for Pantone condition for the first item in the sequence but not for the last one. Indeed, for the last item in the sequence, the difference in recall error between the conditions is much smaller than for the previous items. This suggests that what differentiates the two conditions the most is the involvement in long-term memory. More specifically, Pantone seems to be loaded more heavily than the traditional condition.

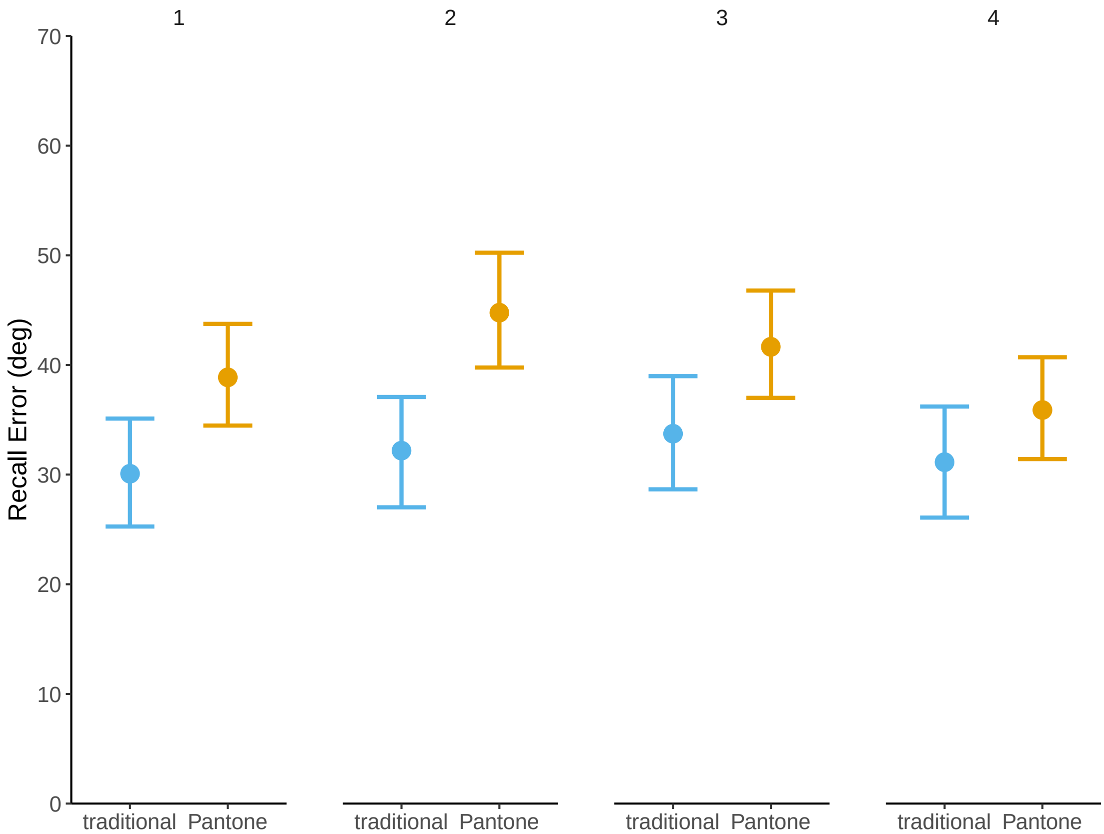

Figure 5: Effect of item position in the presentation sequence on recall error. The smallest difference between the conditions in the last presented item suggests that it is the long-term memory overload that affects the Pantone condition more than the traditional condition.

3.4. Estimating the precision of color memory with deviation from the target color

The next step in the analysis is to test the precision of the remembered colors between both conditions. If Pantone names cannot rely on long-term memory, then maybe we also would not observe the category center bias typically observed for traditional labels [4]. To compare the precision of the remembered color between traditional and Pantone conditions we can look into the deviation of the target color from the chosen one and compare the width of the estimated distribution of responses for each color.

Based on Figure 6 it seems likely that Pantone labels resulted in wider distributions around the target color than traditional labels, however, not necessarily for the color "pink".

The measure of deviation from the target color can also provide us with information on the influence of using different labels for boundary colors in the traditional label condition (see panel B in Fig 6). We hypothesized that a label can pull the remembered color in the direction of the color value prototypical for that label. Thus, for the same color value presented with either the label of the closest color to the left or right we would observe that deviation from the target follows the direction from which the color label was used. The Pantone condition can serve as a neutral label as each boundary color had its unique label.

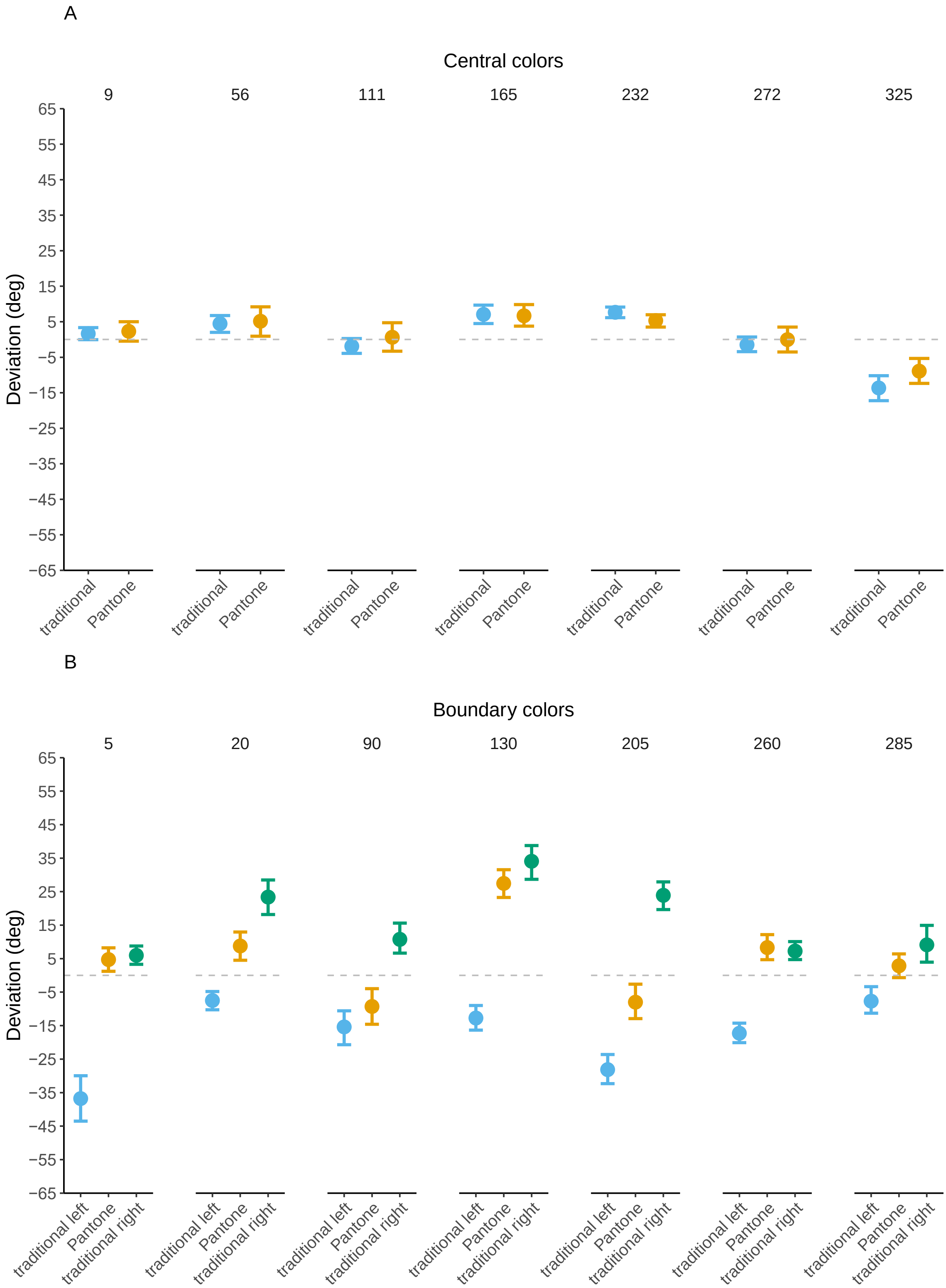

Figure 2: Deviation of the chosen color from the target color for central (panel A) and boundary (panel B) colors. In the traditional label condition, each boundary color had either label borrowed from the central color directly to its left or its right. We observe a systematic effect on the deviation from the target color as a result of manipulating the label the target color was accompanied by. In the traditional condition, if the boundary color was accompanied by the label to "its left" on the color spectrum then the recalled hue was closer to the central color associated with this label and vice versa. Pantone labels for boundary colors were typically the least deviated from the target.

3.5. Estimating the effect of condition order on recall error

We manipulated the order of the traditional and Pantone conditions between participants as we hypothesized that having experience with either Pantone or traditional labels could influence performance in the following condition. However, we did not have directional hypotheses. Figure 7 shows the estimated recall error for both conditions depending on whether the experiment started with Pantone or traditional condition.

The mean recall error is affected by the condition order; however, likely only for the Pantone condition. These results suggest that having experienced the traditional condition made the Pantone condition easier. Through verbal reports at the end of the experiment, we learned that some participants ignored Pantone labels and used traditional ones. This could be one explanation for observing a lower recall error in the Pantone condition after performing the traditional condition.

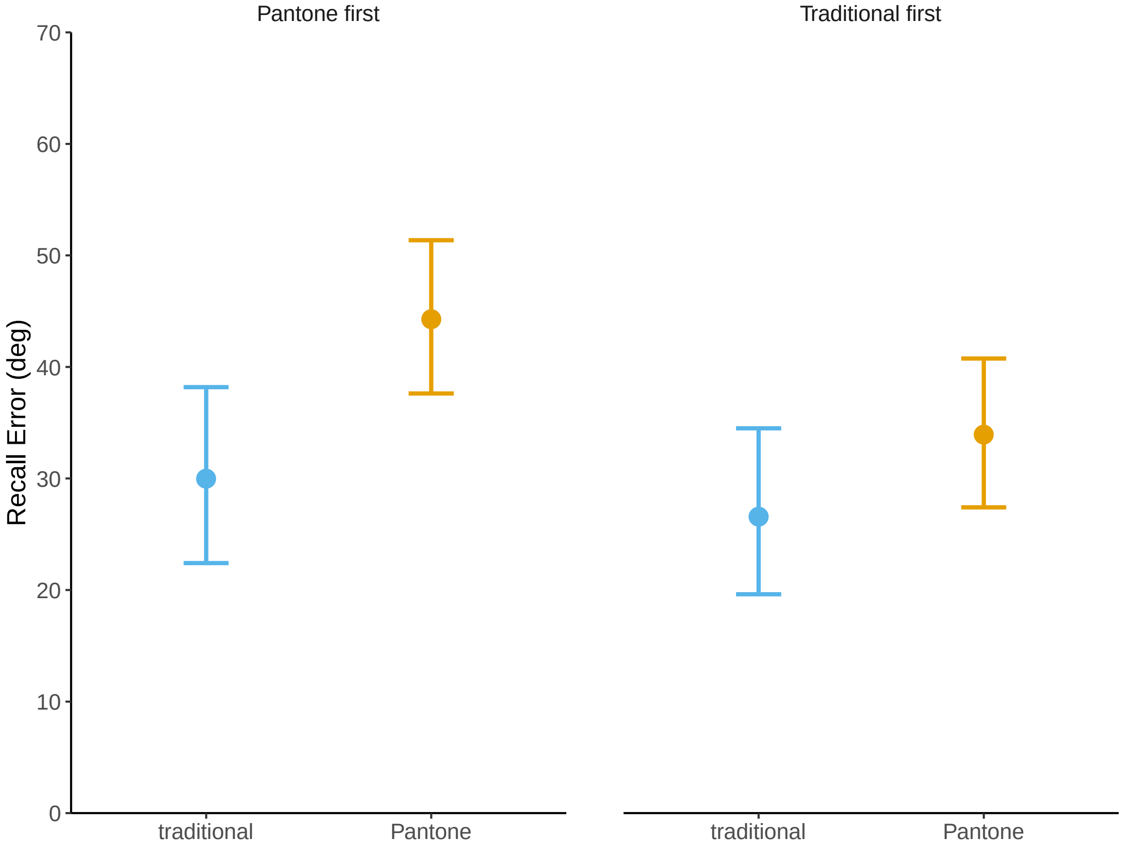

Figure 3: Effect of condition order on recall error

4. Discussion

As in the previous experiments on color memory [12, 4], we also observed that the specific hues we presented during the experiment differ in how well they are remembered. The difference between these experiments and ours is that instead of asking participants to name the colors we played the labels. This suggests that the effects of hearing color labels and labeling yourself may be comparable thus bridging studies focused on the effects of naming on color memory [4, 5] with some new studies on the effects of hearing labels on color perception [7].

Unexpectedly, providing distinct (Pantone) labels for boundary colors did not lead to an improvement in color memory. Moreover, the Pantone labels led to a poorer recall than traditional labels. One explanation of this effect is that Pantone labels are longer than traditional ones, and could lead to overloading the memory, preventing from remembering the exact positions of the colors. As a result, even if the labels improved the precision of the memorized colors, there likely were more errors associated with recalling the correct position the color was presented in. The detrimental effect of long-term memory overload associated with Pantone labels could be supported by the observed primacy and recency effects. The memory of the colors presented at the beginning of the memorizing sequence relies more heavily on long-term memory than colors presented just before the test phase. Based on the recall error associated with each color position we observed that Pantone labels led to poorer performance at the beginning of the sequence than at the end, as compared with traditional labels. This is especially clear for the last item, which instead of long-term memory, is kept in the focus of attention. For the last item, the difference between traditional and Pantone labels is the smallest.

Comparable to previous experiments, we also observed that boundary colors are associated with poorer performance than central colors. This effect is observed for both traditional and Pantone labels. Interestingly, the difference in recall error between boundary and central colors is smaller for the Pantone condition as compared with traditional. A possible explanation is that the more distinctive Pantone labels did lead to a less biased color recollection, at a cost of recollecting fewer colors in general.

To further investigate the bias caused by the traditional labels on boundary colors, we looked into the effects of playing traditional labels "borrowed" from left or right central colors while presenting a boundary color. We observed a clear effect of label-driven color recall bias. When we presented, for example, a boundary color that was represented by a hue between red and orange, if labeled "red" the estimated deviation between the presented hue and the recalled one indicated a higher propensity to recall the color as redder than it was. Alternatively, when the hue was accompanied by the label "orange" the deviation pointed more towards the orange hue. To the best of our knowledge, this is the first demonstration of how label-induced bias can be controlled in color memory studies.

Interestingly, depending on the label (either from the color to the left or to the right) some recalled hues were close to those influenced by the Pantone label. This effect could reflect which of the traditional labels the Pantone label was closest to. For some hues, the Pantone label seemed to provide the smallest bias as reflected by the deviation of the recalled color from the presented target color.

4.1. Study limitations

Pantone labels were not only introducing distinctiveness for boundary colors but also may have overloaded memory performance more than traditional labels. A typical strategy in such tasks is to silently repeat color names to remember the order in which they were presented. The easier the label, the more labels can be maintained in memory. Pantone labels often consisted of two words per hue which may have negatively affected the ability to remember the positions of all the presented hues. In the after-experiment short interview, some participants reported actively ignoring the Pantone labels as they introduced too much information to hold in memory. Some participants also reported replacing the Pantone labels by repeating the traditional labels instead. This could have led to observing a better performance in the Pantone condition when it was preceded by the traditional condition.

5. Conclusions and future directions

Overall, based on the reported study we have some evidence in favor of more distinctive labels improving the precision of color memory. However, this study should be considered as an opening of a research line focused on fine-tuning the procedural details and expanding the applicability of the results, possibly to professionals working daily with colors. What first needs to be tackled is the choice of appropriate novel labels which would be more distinctive than the seven traditional ones but at the same time would not cause additional strain on memory.

Additionally, to assess whether Pantone labels affected the recall of the correct hue based on forgetting its position, formal modeling can be performed using a mixture model developed in the working memory research area that allows assessing the rate of recalling the correct hue but at the wrong position (so-called "swap errors" [13]).

To remove the detrimental effect of memory overload due to too-long labels in Pantone a follow-up study could be run in which unique one-word labels for boundary colors could be included. This could require an additional pilot study to collect verbal labels from participants regarding both central and boundary colors. Crucially, the labels for boundary colors would have to be provided with achieving their distinctiveness from central colors in mind. Otherwise, the results would likely be comparable with labeling behavior observed in a previous study assessing spontaneous labeling [5].

6. References

[1] Berlin, B., & Kay, P. (1991). Basic color terms: Their universality and evolution. Univ of California Press.

[2] Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. Proceedings of the National Academy of Sciences, 104(19), 7780-7785.

[3] Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. Journal of Experimental Psychology: General, 144(4), 744.

[4] Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. Cognition, 166, 277-297.

[5] Souza, A. S., Overkott, C., & Matyja, M. (2021). Categorical distinctiveness constrains the labeling benefit in visual working memory. Journal of Memory and Language, 119, 104242.

[6] Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. Journal of Vision, 13(7), 1-1.

[7] Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. Journal of Experimental Psychology: General, 148(7), 1105.

[8] Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. Nature, 453(7192), 233-235.

[9] Zhou, K., Mo, L., Kay, P., Kwok, V. P., Ip, T. N., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. Proceedings of the National Academy of Sciences, 107(22), 9974-9978.

[10] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[11] Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software, 80, 1-28.

[12] Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. Journal of Vision, 14(4), 7-7.

[13] Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. Journal of Vision, 9(10), 7-7.