Color Naming in Two Languages

CSE 512 Final Project Report

Kyle Thayer
kthayer@cs.washington.edu

Younghoon Kim
yhkim01@cs.washington.edu

Gabriella Gorsky
gsg13@uw.edu

ABSTRACT
In data visualization, one of the most critical considerations for designers is that of color. Previous research has shown that language can affect how colors are named and perceived in specific instances [25]. While some work has been done on collecting color names in English [18] and unwritten languages [13], many common languages have not had their color naming formally cataloged. We ran a pilot study for collecting color names in different languages and collected sufficient data for a preliminary analysis of English and Korean. We found differences in how these languages name colors, particularly in the blues, and we lay the groundwork for collecting a larger range of colors and languages in the future.

INTRODUCTION
In data visualization, one of the most critical considerations for designers is that of color. Color choices can have a profound impact on how effectively information is communicated to a particular audience. Early research on the psychology of color perception explored the relationship between colors and mental states [10]. Interest in the relationship of color to affective states and other individual personality and individual characteristics continued to play an important role in psychological research on color perception [12], with increasing attention paid to the role of language and culture on this behavior [2]. This foundational work in psychology paved the way for contemporary color perception studies exploring color preferences among different cultures [15], with an increasing interest in patterns of color naming and association within and across different cultures [14]. Research studies conducted over the last two decades have revealed language differences in color naming behavior [25, 3, 4, 8, 26, 13].

Previous work has been done to graphically model colors according to how people name them. However, these efforts are mostly limited to English language speakers [18, 9, 6], with the exception of the World Color Survey which collected color naming data on 110 unwritten languages [13]. In order to further explore these differences, we designed a pilot study with the goal of collecting color naming data from a broader range of languages than in previous studies.

We aim to explore these differences by collecting color naming data from a broader range of languages. To gather these data, we ran a short-term pilot study on LabintheWild (www.labinthewild.org), which hosts behavioral research studies using voluntary web sampling in an attempt to reach a larger-scale and broader audience, as well as to propagate more reliable and higher quality data. The LabintheWild platform attempts to address potential demographic limitations by attracting a more diverse range of participants through viral social media approaches [23].

This work seeks to contribute the following:

- Extend previous work on color naming in different languages
- Utilize the LabintheWild survey platform to reach a broad audience, but without some of the drawbacks of Mechanical Turk administration
- Better incentivize participation in the survey through the inclusion of dynamic color tasks
- Collect reliable data that will reveal differences in color-naming between languages

RELATED WORK
Prior work shows that language can have dynamic effects on color perception. In Russian, for example, there are two high-level words for blues: light blue (”*goluboy*”) and dark blue (”*siniy*”). In contrast, English has a single high-level word for the same set of hues: ”blue.” One study of Russian speakers found that they were faster at discriminating two distinct colors when they came from separate categories (one *goluboy* and one *siniy*) than if the colors came from the same blue category (both *goluboy* or both *siniy*). No speed difference was found across conditions for English speakers [25]. Similar effects were demonstrated in studies involving speakers of Uruguayan Spanish, Greek, and Japanese, which are all languages in which a similar light-blue/dark-blue distinction exists [3, 8, 4].

In addition to the relationship between naming and identification, other research has explored the interaction between visual perceptual differences related to language and color naming. Mandarin-English speakers shown color names in Mandarin and English and asked to select the best name for a color patch took longer when choosing between Mandarin names if the color patch was displayed in the right visual field. This phenomenon can be explained by the fact that Mandarin script is read from right to left [26]. Even in spoken languages without written script, speakers of each language exhibit differences in segmenting color space [8, 7].

In order to extend previous research on color naming and employ a more reliable color survey, several considerations were made regarding research design and survey delivery. The exploratory study which serves as the basis for the present work utilized a Mechanical Turk crowd-sourcing approach. However, several limitations to this data collection approach muddle overall interpretations of the findings. In general, Mechanical Turk's demographics overwhelmingly skew towards English speaking workers in the US and India. In addition, because workers are paid for task completion, some workers try to game the system by completing tasks as quickly as possible with little or no attention paid to the accuracy of their responses [11, 21]. Coupled with the fact that color naming responses were given in languages that did not match the native language of the Turker, it was necessary to utilize LabintheWild as an alternative mode of color survey delivery and respondent sampling.

One critical, but often overlooked, component of online color naming research relates to the effect of situational lighting conditions on participants ability to see color. Given that color perception can already be highly variable across individuals, it is relevant to consider the findings of new research in this area that found that there is a significant effect of room and monitor brightness on color discrimination, with increased task difficulty as lighting ratio increased [22]. In order to account for this, the current study asks participants about the lighting conditions of the setting in which they are completing the color tasks which will serve to counterbalance potential confounds.

Currently, much of the relevant recent research on modeling color naming data focuses on English color words, but nonetheless provides a solid foundation for deeper exploration of differences in naming patterns across languages [6, 18, 9]. The present study applies similar non-parametric probabilistic modeling approaches to a cross-lingual data set in order to identify differences in the perceptual mapping of color. It is relevant to acknowledge much of the existing work on gamut mapping uses color naming data spanning the entire three-dimensional color space [16, 17], whereas the present study focused only on L*a*b-scaled pure hue values. Future study iterations will expand the color space to account for full HSV color space.

DATA COLLECTION

We carried out our data collection through a pilot experiment on LabintheWild [23], an online experiment platform where the users are not motivated to participate in the experiment financially, but instead are motivated to participate and share the experiment because they can learn something about themselves and compare themselves to their peers.

The experiment we piloted on LabintheWild for this study had 3 sections–demographics questions, two types of tasks for the participantss (color naming and color sorting), and a results page where the user receives a color vision score that they can share with their friends on social media. The experiment had 5 stages: 1, 3, and 5 were color naming, 2 and 4 were color sorting, and there was a break between stages 3 and 4 that had a photo of a kitten and a note encouraging users to rest their eyes before continuing.

While the color naming task produced the primary data we were interested in collecting, the color sorting task gave us a number of benefits. It provides us with additional data on the color vision and discrimination abilities of our users. We also believe it may encourage users to think more about subtle color differences while they are naming. Most importantly, it allows us to give the participants information about themselves that will hopefully motivate them to take the test and share it with others.

Below we describe each piece of the study, as well as the motivation behind the design.

Demographics

After the initial consent page of our study, the participants are shown a set of demographics questions. These questions include some standard questions about countries lived in, native language and other languages known, gender, age, and education, along with questions on monitor brightness and external lighting [22], color vision deficiency, and whether the participant's education or work background includes working with color names.

Color Naming Task

Our experiment had three pages of color naming tasks (see Figure 1). For these we asked the user to name colors in the language they indicated as their native language in the demographic survey. We asked them to use the most common character set for their language because in our previous work we found users entering color names in different scripts, which made data cleaning more difficult.

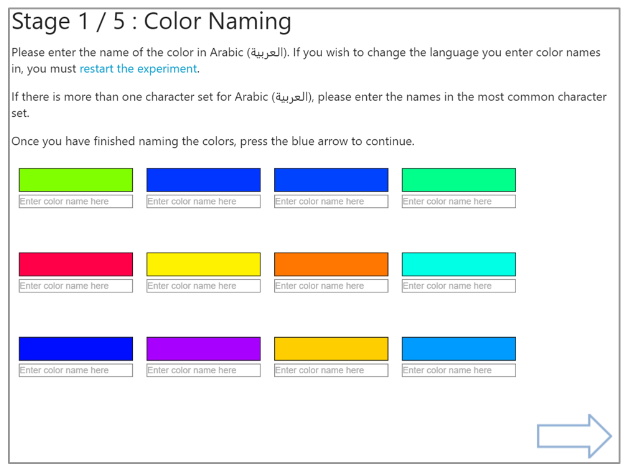

Figure 1. Screen shot of the color naming task.

For each color naming stage the user is shown 12 color swatches (for a total of 36 swatches seen throughout the whole experiment). Each tile is 150 pixels wide and 30 pixels high and has a 0.5 pixel black border around it. Below the tile is a text box where users can enter the color name. For Chinese and Korean we included code that detected whether the user was entering in the normal script and included a note requesting the normal script if they entered text in a different script (eg. if they started typing in characters from the Latin alphabet).

The colors the users were asked to name came from a path along the edges of the RGB cube with full saturation and highest brightness. This works out to be the path between the following RGB values: (255, 0, 0), (255, 255, 0), (0, 255, 0), (0, 255, 255), (0, 0, 255), (255, 0, 255), and back to (255, 0, 0). We chose to restrict ourselves to a one dimensional set of colors to reduce the amount of data needed to get results, and we chose this particular path through the data based on exploring the XKCD color dataset [18] and looking at which paths had the most clearly named colors in English.

Once we had our path of RGB colors, we converted them into the more perceptually driven L*a*b color space [24] and calculated the distance between each consecutive pair of colors along the path. We used these distances to re-scale the colors and divided them into 36 bins of equal length in L*a*b. Each participant of the study was shown one color, randomly chosen from each of these 36 bins. The bins guaranteed that each participant was asked about a representative set of colors along our path, which supported later binning, and the random locations within the bins could support different binning if enough data was collected.

Color Sorting Task

Our experiment included two pages of color sorting tasks (see Figure 2). Our task is similar to the The Farnsworth-Munsell 100-hue and dichotomous tests for color vision [20] which involves sorting 100 physical color tiles, and their online "Color IQ" version which involves sorting 80 colors [1].

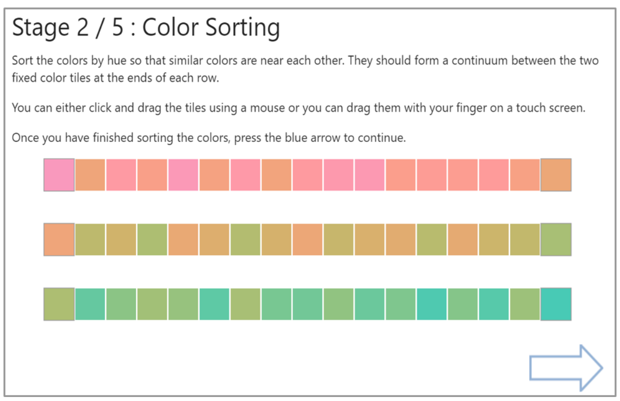

Figure 2. Screen shot of the color sorting task.

Unlike the Farnsworth-Munsell 100-hue test, which chooses colors from the older, perceptually-based Munsell color space [19], we chose the more modern perceptually-based L*a*b color space [24]. From this space we chose 90 colors from a circle centered in the a*b* plane of uniform L* (lightness) value. We chose 90 since it was close to the 100 from the Farnsworth-Munsell 100-hue test, but it divided into even 4 degree angle increments. We sought the L* value that would allow us the largest radius in order to have the most saturated version of colors. We found the best such radius, with up to two decimal places, was an L* value of 74.04 and a radius of 40.26.

We took these 90 colors and divided them into 6 ranges. The Farnsworth-Munsell 100-hue test divides them into 4, but we felt this made each row of tiles too wide and burdensome for our users. We presented each participant with the 6 ranges in hue order, the first 3 in stage 2 and the second 3 in stage 4.

For the sorting task, the first and last tiles were anchored, with a given anchor coming from the appropriate first or last colors of the adjacent ranges. Participants could drag and drop each of the interim 15 tiles in each range to sort them into the correct order. This demonstrated their color discrimination ability and primed them for subsequent color-naming tasks.

Results Page

After each participant completed the 5 stages, they were given an opportunity to provide feedback and then shown their results page (see Figure 3). The results page gave the user their color vision score, told them how they compared to the average score and showed them their color perception spectrum.

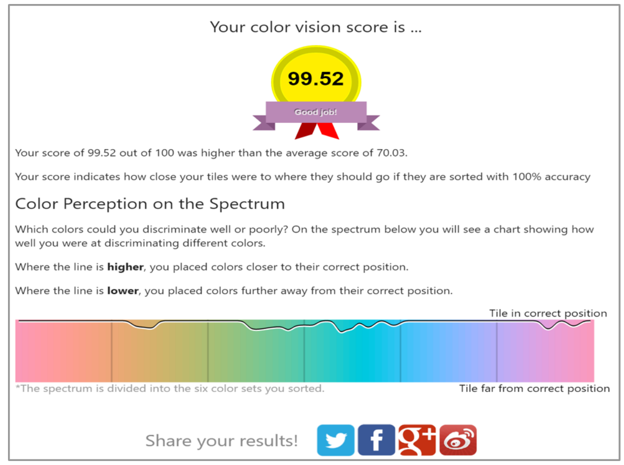

Figure 3. Screen shot of the results page.

The color score was calculated by computing an error as the sum of the squares of each tiles distance from its correct position. We chose to use the square to increase the penalty of larger errors in placement. We then found the maximum possible error (perfect sorting of the tiles backwards) and then set their score to 1 - (participant error / maximum error). This scoring system ended up giving anyone who made a decent effort very high sores (the average score was 98.89%).

For the spectrum, we set the height of the line as the ratio of the participant's distance from the correct position for a tile and the farthest off that tile could be. This then has the line at the top if the participant placed the line in the correct position, and any dips indicate where a user misplaced tiles and by how much.

At the bottom of the page were links to common social networking sites such as Twitter and Facebook where users could share their score and compare scores with any friends who were willing to share theirs.

Recruitment

In order to recruit participants, we posted links to the study on Facebook under our own profiles and also the official LabintheWild Facebook page. We additionally encouraged friends and family to take the test and share their results on social media such as Facebook and Twitter. We did not put the experiment on the homepage of LabintheWild since this was a pilot of the experiment.

RESULTS

We conducted the survey for 5 days and obtained 10,428 color names from 347 participants in 14 different languages (Table 1). However, only English and Korean provided sufficient name data to be analyzed. As a result, we restricted our analysis to a comparison between those two languages.

Language	# of Responses	# of Participants
English	8832	293
Korean	864	32
Polish	156	5
German	108	3
Spanish	96	3
Portuguese	72	2
Russian	72	2
Chinese	36	1
French	36	1
Hebrew	36	1
Japanese	36	1
Norwegian Bokmål	36	1
Thai	36	1
Urdu	12	1
Total	10428	347

Table 1. The number of responses and participants in each language. Only English and Korean were gathered enough to be analyzed.

Of our sample, 48% of our English participants and 75% of Korean participants reported they've completed or are in graduate school. Gender ratios (percentages of male/female/other) for each group were 26%, 72%, 2% for English and 59%, 41%, 0% for Korean, with women over-represented in the English subset and men slightly over-represented in the Korean subset.

Exploratory Statistics

In order to analyze the responses quantitatively, we applied the color naming model [9] to these data with an additional variable: language. From this model, we mainly focused on generating the probability of a color name, given a particular color swatch [P(n|c)]. We used (c,t,l), given c rather than given t, because it was easier to see what terms dominated each binned color. Based on this metric, we examined to what degree Korean and English speakers name colors differently, and made an interactive visualization to show the two distributions (screen shot in Figure 5).

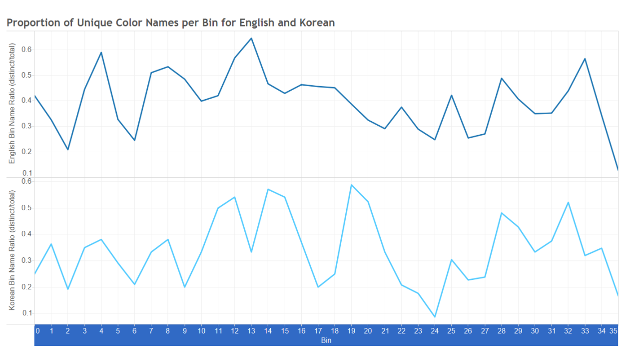

Figure 4. Line chart showing the ratio of distinct names given to a color bin out of total names applied in English and Korean.

In addition to calculating the probability of a color name given a particular color bin, we were interested in understanding if those trends could be further explained by differences in overall naming conventions. In order to explore this, the ratio of distinct names out of total names given to a color bin were calculated for Korean and a series of randomly sampled subsets of English, due to the sampling imbalance between the two languages (single group N = 803).

Paired-samples analyses revealed a significant difference between the English and Korean ratios ($t(35) = 2.67, p < .05$). English tended to have greater proportions of distinct words/total words applied to a bin. This is consistent with probabilities, which revealed less naming agreement in English than Korean, especially for light green and light blue (evident for bins 13 and 17 in Figure 4).

Further exploration of a more complete dataset will allow for increasingly robust conclusions regarding significant trends in color naming. In addition to more reliable estimates of distinct color name values, Kullback-Leibler Divergence (also known as relative entropy) will be applied as a measure of the difference between the probability distributions of the two languages.

Visualization

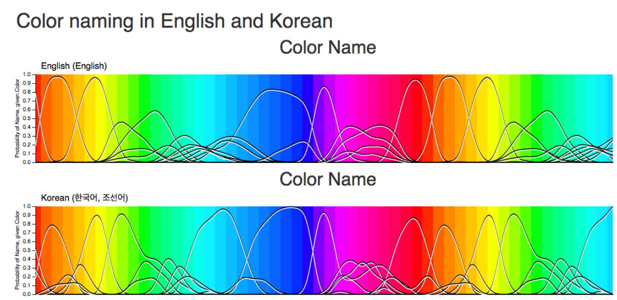

Figure 5. Screen shot of visualization. Each line present the probabilities of a color name, given a particular color swatch.

We visualized the data as multi-line charts in two separate rows to present the differences of color naming in two languages, using D3 [5] (Figure 5). In preliminary research [9] and in a previous project (a3-kylethayer-yhoonkim), the researchers tried to express the color space through projection into 2D plane. On top of the positional variables, they mapped external fields, such as the degree of agreement, to other visual variables like size. But in our study, we narrowed down the number of colors as a closed path following edges of RGB color region on the L*a*b space. Thus, we were able to present the colors on the x-axis and use the y-axis for the probabilities.

We also repeated half of the color spectrum to prevent the distributions from being disconnected at two ends of the x-axis. We varied background color according to the colors on x-axis, and for the sake of being contrasted from those background color, we composed each line with black and white color. Finally, we interpolated the lines by 'basis' feature provided by D3. This make the distributions more smoothly but harmed to read peak values.

With regard to comparison between two languages, we plotted the same style charts in parallel. We allowed users to mouse over on each line to highlight it and see its color name, and to click it to hold the highlight and the name. Due to the fact that lines are too thin to be hovered by mouse cursor, we put thick transparent line on the top of each line, and attached the event trigger on them. In addition, we provide translations of each color name to help to understand what the words mean.

DISCUSSION

Overall, the LabintheWild pilot survey allowed for a more nuanced exploration of color naming patterns in different languages than the previous study using Mechanical Turk. In addition to providing more demographic information from participants, which can be used for quantitative comparisons in future work, the approach generated more consistent data due to the more self-directed motivation framework.

While data was limited for languages other than English, a satisfactory number of Korean responses allowed for an exploratory analysis of naming trends. While the Korean data was relatively sparse, it suggests some interesting patterns that can be explored in greater detail following a second, longer data collection period.

The visualization we produced effectively revealed fascinating color naming differences between English and Korean in a visually appealing and intuitive way. It allows viewers to see naming differences across the continuous color spectrum. The most interesting finding was that Korean has a distinct color name for light green and light blue, while English does not. It is unclear exactly why this is the case, but it is consistent with and extends related research on languages with similar light-blue/dark-blue distinctions [3, 8, 4]. These findings suggest that research exploring how Korean language-users also utilize this two-hue schema is warranted.

Limitations

There were a number of limitations to the design and implementation of this study. The decision to choose a limited color space was necessary in order to maximize the amount of per-color data we could collect during the short release of the survey. However, the fact that colors were sampled from only some edges of the RGB cube means that we can get only a partial picture of how language interacts with color perception and naming.

There were also some significant limitations to our analysis. While we were able to gather a satisfactory total number of study participants, we weren't able to get broad language coverage. Responses were only dense enough in English and Korean. Even then, English was drastically over-represented (10:1) which made for less reliable comparisons across the two languages. In addition to language imbalance, as a result of employing a convenience sampling approach to online data collection, more than half of the participants have been or are in graduate school. Further, study instructions were only administered in English, which means that all participants had to have some level of proficiency in English in order to even complete the survey, despite a research interest centered on linguistic salience. These factors present some clear implications for any claims we might make about trends in the data.

Future Work

While this pilot study allowed us to do some preliminary data exploration, it also provided many opportunities for refining and improving our approach in the subsequent survey release. Collecting a much broader range and density of data will be imperative to drawing more reliable conclusions about color naming trends across languages. In addition, translation of the survey into a few focal languages will allow for the collection of more culturally salient responses which may better reflect the relationships between language and color perception.

Other considerations for future work include systematizing the handling of spelling and structural differences in color naming (ie. ”yellowish green” vs. ”yellow green” or ”fuchsia” vs. ”fuchia”) which may obscure or overemphasize color naming differences. In addition, some participants reported that they had minor eye strain during the sorting tasks in spite of the break page in the middle of the tasks. Moreover, participants tended to spend less time on the naming tasks. Therefore, it will be a good next step for collecting color names better to balance the loads between two types of tasks.

CONCLUSION

Through the LabintheWild pilot study, we furthered our understanding of how language and culture interact with visual perception in English and Korean. Continued exploration will allow us to address some of the limitations to the pilot and produce data that can be used to more directly investigate how color perception varies across multiple languages, providing valuable insight into effective visual design choices for diverse audiences.

REFERENCES

1. Test Your Color IQ: X-Rite Photo & Video. (????). `http://xritephoto.com/cool-tools`
2. Francis Adams and Charles Osgood. 1973. A Cross-Cultural Study of the Affective Meanings of Color. *Journal of Cross-Cultural Psychology* 4, 2 (1973), 135–156.
3. Panos Athanasopoulos. 2009. Cognitive representation of colour in bilinguals: the case of Greek blues. *Bilingualism: Language and Cognition* 12, 1 (1 2009), 83–95. DOI: `http://dx.doi.org/10.1017/S136672890800388X`
4. Panos Athanasopoulos, Ljubica Damjanovic, Andrea Krajciova, and Miho Sasaki. 2011. Representation of colour concepts in bilingual cognition: the case of Japanese blues. *Bilingualism: Language and Cognition* 14, 1 (1 2011), 9–17. DOI: `http://dx.doi.org/10.1017/S1366728909990046`
5. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3: Data-Driven Documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* (2011). `http://vis.stanford.edu/papers/d3`
6. Jason Chuang, Maureen Stone, and Pat Hanrahan. 2008. A Probabilistic Model of the Categorical Association between Colors. In *Color Imaging Conference*. 6–11. `http://idl.cs.washington.edu/papers/color-names`
7. Richard S. Cook, Paul Kay, and Terry Regier. 2005. Exploring the Opportunities and Challenges with Exercise Technologies for People who are Blind or Low-Vision. In *Handbook of Categorisation in the Cognitive Sciences*. ACM. `http://www1.icsi.berkeley.edu/~kay/claire7.pdf`

8. Antoni Gomila. 2012. Language as Lens: Lexical Differences. In *Verbal Minds: Language and the Architecture of Cognition*, Elsevier (Ed.).

9. Jeffrey Heer and Maureen Stone. 2012. Color Naming Models for Color Selection, Image Editing and Palette Design. In *ACM Human Factors in Computing Systems (CHI)*. http://idl.cs.washington.edu/papers/color-naming-models

10. K. Hevner. 1935. Experimental Studies of the Affective Value of Colors and Lines. *Journal of Applied Psychology* 19, 4 (1935), 385–398.

11. Panagiotis G. Ipeirotis. 2010. Analyzing the Amazon Mechanical Turk Marketplace. *XRDS* 17, 2 (Dec. 2010), 16–21. DOI:http://dx.doi.org/10.1145/1869086.1869094

12. Keith W. Jacobs and James F. Suess. 1975. Effects of Four Psychological Primary Colors on Anxiety State. *Perceptual and Motor Skills* 41 (1975), 207–210.

13. Paul Kay and Richard S. Cook. 2015. World Color Survey. In *Encyclopedia of Color Science and Technology*, Ronnier Luo (Ed.). Springer Berlin Heidelberg, 1–8. http://link.springer.com/referenceworkentry/10.1007/978-3-642-27851-8_113-10 DOI: 10.1007/978-3-642-27851-8_113-10.

14. Timothy D. King. 2005. Human Color Perception, Cognition, and Culture: Why Red is Always Red. *Color Imaging X: Processing, Hardcopy, and Applications* 5667 (2005), 234–242. Issue 1.

15. Thomas J. Madden, Kelly Hewett, and Martin S. Roth. 2000. Managing Images In Different Cultures: A Cross-National Study of Color Meanings and Preferences. *Journal of International Marketing* 8 (2000), 90–107. Issue 4.

16. Hideto Motomura. 2001a. Categorical Color Mapping Using Color-Categorical Weighting Method–Part I: Theory. *Journal of Imaging Science and Technology* 45 (2001), 117–129. Issue 2.

17. Hideto Motomura. 2001b. Categorical Color Mapping Using Color-Categorical Weighting Method–Part II: Experiment. *Journal of Imaging Science and Technology* 45 (2001), 130–140. Issue 2.

18. Randall Munroe. 2010. XKCD Color Survey Results. (2010). http://blog.xkcd.com/2010/05/03/color-survey-results/

19. Albert Henry Munsell. 1919. *A color notation*. Munsell color company.

20. Dorothy Nickerson, Walter C Granville, and J Opt. 1943. The Farnsworth-Munsell 100-hue and dichotomous tests for color vision. *J. opt* (1943).

21. Ellie Pavlick, Matt Post, Ann Irvine, Dmitry Kachaev, and Chris Callison-Burch. 2014. The Language Demographics of Amazon Mechanical Turk. *Transactions of the Association for Computational Linguistics* 2 (2014), 79–92. https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/262

22. Katharina Reinecke, David R. Flatla, and Christopher Brooks. 2016. Enabling Designers to Foresee Which Colors Users Cannot See. ACM Press, 2693–2704. DOI: http://dx.doi.org/10.1145/2858036.2858077

23. Katharina Reinecke and Krzysztof Z. Gajos. 2015. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 1364–1378. DOI:http://dx.doi.org/10.1145/2675133.2675246

24. Alan R Robertson. 1977. The CIE 1976 Color-Difference Formulae. *Color Research & Application* 2, 1 (1977), 7–11.

25. Jonathan Winawer and others. 2007. Russian blues reveal effects of language on color discrimination. In *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences. http://www.pnas.org/content/104/19/7780.full

26. Sophie M. Wuerger and others. 2012. Blue–green color categorization in Mandarin–English speakers. *Journal of the Optical Society of America* 29, 2 (2 2012), 102–107. DOI:http://dx.doi.org/10.1364/JOSAA.29.00A102