Color Names Across Languages: Salient Colors and Term Translation in Multilingual Color Naming Models

Younghoon Kim, Kyle Thayer, Gabriella Silva Gorsky and Jeffrey Heer

University of Washington

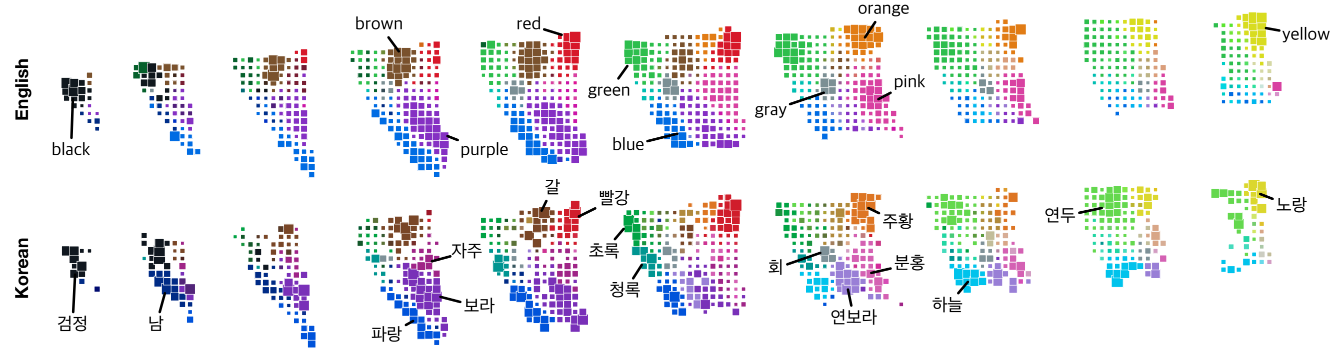

Figure 1: *Maximum probability maps of English and Korean color terms. Each point represents a* $10 \times 10 \times 10$ *bin in CIELAB color space. Larger points have a greater likelihood of agreement on a single term. Each bin is colored using the average color of the most probable name term. Bins with insufficient data (*< 4 *terms) are left blank. English has 10 clusters corresponding to basic English color terms [BK69], whereas Korean exhibits additional clusters for 남 ( ), 청록 ( ), 자주 ( ), 하늘 ( ), 연두 ( ), and 연보라 ( ).*

Abstract

Color names facilitate the identification and communication of colors, but may vary across languages. We contribute a set of human color name judgments across 14 common written languages and build probabilistic models that find different sets of nameable (salient) colors across languages. For example, we observe that unlike English and Chinese, Russian and Korean have more than one nameable blue color among fully-saturated RGB colors. In addition, we extend these probabilistic models to translate color terms from one language to another via a shared perceptual color space. We compare Korean-English translations from our model to those from online translation tools and find that our method better preserves perceptual similarity of the colors corresponding to the source and target terms. We conclude with implications for visualization and future research.

CCS Concepts

• ***Human-centered computing*** → ***Visualization design and evaluation methods;*** *Visualization systems and tools;*

1. Introduction

Associations between colors and linguistic terms (names) are valuable to consider when choosing colors for visual communication. *Salient* colors with unambiguous names make it easier to refer to, recognize, and recall graphical elements [RDD00]. Prior work contributes color name models in English as well as for unwritten languages [CSH08, HS12, KBM*09]. Though specific instances of naming and perceptual differences have been studied across languages [GPRMA17, Ath09, ADKS11, WWF*07], we lack multilingual color naming models to aid visualization and graphic design.

We construct color naming models [CSH08] across languages based on crowdsourced color name judgments. We contribute:

- Two datasets of human color-name judgments: names for *saturated hues* (with full saturation and brightness in HSV space) across 14 common languages, and names for *full colors* (the entire RGB cube) in English and Korean.
- Probabilistic color naming models for each language and color set. For the saturated hues, we investigate how the most nameable (salient) colors vary per language. For example, we confirm that Korean and Russian have two nameable blue colors (하늘 , and 파랑 , **голубой** and **синий**) in the saturated hues [WWF*07], unlike other languages (*e.g.,* blue in English).
- Application of our multilingual models to perform color term translation. We compare our results to translations from popular English-Korean online translation tools. We discover question-

© 2019 The Author(s)
Eurographics Proceedings © 2019 The Eurographics Association.

able translations by the online tools, such as 자주 (■) for purple (■), whereas our method maintains perceptual fidelity.

Our collected data and color naming models are available online at https://github.com/uwdata/color-naming-in-different-languages.

2. Related Work

The commonality of color names across languages has been studied and debated. Universalists argue for innate color perception mechanisms, while relativists point to variations in color terms across languages and cultures [KR06]. To investigate this issue, Berlin and Kay first collected color names across 20 languages from one bilingual speaker per language in the San Francisco Bay area [BK69]. Their World Color Survey [KBM*09] then collected color-name pairs for 110 unwritten languages. In this work, we collect human color-name judgments from multiple native speakers of common written languages using LabintheWild [RG15], an online platform for crowdsourced experimentation.

To model associations between colors and names, Heer & Stone [HS12] apply Chuang et al.'s probabilistic modeling approach [CSH08] to over 3M English color name judgments gathered from readers of the web comic XKCD [Mun10]. This model allows calculation of a color's saliency (the degree of naming consensus for a given color), and replicates identification of Berlin & Kay's basic color terms [BK69]. Heer & Stone then apply the model to evaluate color palettes for data visualization. We build similar models for multiple languages, and use them to analyze and compare color saliency across languages.

3. Data Collection

We collected color name judgements for different languages in an experiment on LabintheWild [RG15], an online platform with a globally diverse user base. Experiments on LabintheWild motivate users to participate and share the experiment by allowing them to learn something about themselves and compare themselves to their peers. Our experiment advertised that participants could learn about their color perception abilities and compare their results with others. The study can be taken at https://labinthewild.org/studies/color_perception/.

The experiment had three sections: demographic questions (*e.g.,* native language), two color perception tasks (color naming and sorting), and a results page. The naming task produced the data for this study, while the sorting task let us calculate a color perception score for the results page, so that participants could compare their score with the average and share it on social media.

3.1. Task

Our experiment had three pages where we asked the participant to name colors. We asked subjects to give names in their native language, using the most common character set for that language.

For each color naming stage, a participant was shown 12 color tiles (for a total of 36 tiles across the three pages). Each tile was a 150×30 pixel rectangle with a 0.5 pixel black border and white background. Below the tile was a text box where subjects could enter the color name. For Chinese and Korean, we detected if the subject used the expected character set, so we could prompt them to do so if needed.

The stimuli colors came from one of two sets: *saturated hues*, a path along the edge of the HSV color wheel with full saturation ($S = 1$) and brightness ($V = 1$), and *full colors*, colors from the full RGB cube. Initially, we restricted ourselves to the saturated hues to make data collection more feasible. We chose the hue colors in particular because they are commonly used in color pickers (*e.g.,* default pickers in Windows and Mac OS) and we believed these colors would be more straightforward to name. Once a language received 1,000 color-name pairs we began collecting colors from the full RGB cube for that language.

To ensure that each participant is given an approximately perceptually uniform set of colors, we discretize the hue circle into 36 equally-spaced 36 bins within CIELAB color space. Every subject saw one color from each of these 36 bins, with the specific color stimuli randomly sampled from each bin. To sample the RGB cube, we select 36 random colors from the full space, subject to the constraint that all samples must be at least 20 units apart in CIELAB space to ensure that reasonably different colors are presented.

For the color sorting task, we present 90 color tiles to sort. Subjects sorted 15 tiles at a time, and asked to form a smooth gradation between anchored starting and ending tiles. This task was inspired by the Farnsworth-Munsell [Far43] 100-hue and dichotomous tests for color vision, which involve sorting 100 physical color tiles from the perceptually-based Munsell color space. In our test we chose 90 colors (instead of 100) that were equally spaced along the largest centered circle in the a*b*-plane of CIELAB.

3.2. Recruitment

To recruit subjects, we posted links to the study on Facebook and Twitter under our own profiles and the official LabintheWild Facebook page. We also encouraged friends and family to take the test and share their results on social media. To promote more participation, we translated the experiment instructions into Korean, Chinese, and Farsi after launching. One later advertising post included Korean and English names of some hue colors, which may have primed the subjects' color naming, so we exclude hue color-name pairs collected after this posting.

In total we collected 131k color-term pairs from 4.2k participants across 70 languages from May 27, 2016 to February 1, 2019. In this paper we focus on the 14 languages that had at least 500 respondents: English, Korean, German, Spanish, French, Portuguese, Swedish, Polish, Russian, Chinese, Persian, Dutch, Finnish, and Romanian for the saturated hues. For the full colors, we examine Korean and English, the two languages for which we have sufficient coverage. Data collection remains ongoing.

3.3. Data Processing

The free text responses result in a variety of color names. To manage variations in punctuation, we remove non-alphabetic characters (*e.g.,* dashes, underscores, and whitespace). We change all uppercase letters to lowercase when available. For Russian, Korean, Chinese, Arabic, and Persian, we filter out responses with non-native script characters. We also exclude color terms that do not

© 2019 The Author(s)
Eurographics Proceedings © 2019 The Eurographics Association.

belong to the participant's self-reported language by checking if they are listed as colors in native dictionaries. Lastly, we apply customized rules for Korean, English, Persian, Portuguese, Chinese, and French terms that correct typos and merge words that were grammatically the same. The rules were reviewed by at least one native speaker per language (details in supplemental materials).

To aggregate responses, we bin the saturated hues into the 36 sampling bins mentioned earlier and bin the full RGB cube into $10 \times 10 \times 10$ bins in CIELAB space. We omit RGB cube bins with less than 4 observations. Following Heer & Stone [HS12] we also remove sparsely occurring terms to reduce noise: for saturated hues we retain the top 20 terms for each language, and for the full colors we exclude terms that occur only once. This exclusion drops relatively little information. The Frobenius norm of the color-by-name matrix (columns are terms, rows are color bins) is reduced by less than 4% for saturated hues (except for Chinese, reduced by 8%), and less than 1% for full colors.

4. Analysis & Results

The final data set we used to build our probabilistic models had 37,763 saturated hue and 97,785 full color records. We now compare the color name compositions of the saturated hues and the full colors across the languages [CSH08]. We then introduce a translation model by extending the full color models.

4.1. Color Name Compositions Across Languages

To compare agreement for color names, we first compute the probability $P(T|c)$ of terms T within a color bin c. To quantify the degree of naming consensus for a given color, we compute two measures: the negative entropy $saliency(c) = -H(P(T|c))$, and the maximum probability $maxProb(c) = \max(P(T|c))$. Prior work uses the former measure to measure color nameability [CSH08,HS12]. We also employ the latter measure as it indicates the maximum likelihood of agreement on a single term.

In Figure 2, we visualize the naming of hue colors for 14 languages (see supplemental materials for interactive visualizations). The figure shows a common pattern in saliency: 7 color regions corresponding to Berlin & Kay's basic color terms [BK69]: red, orange, yellow, green, blue, purple, and pink (weakly), are salient across languages. However, we also see that some languages have additional salient colors: Korean has two nameable blue colors (하늘 and 파랑), as does Russian (the previously studied **голубой** and **синий** [WWF*07]). In addition, we observe a salient Russian teal color (**бирюзовый**).

In comparison to the other languages, Chinese color name saliency was relatively low, exhibiting higher naming variation. This observation may explain why the Chinese dataset saw a greater reduction (%8) than the other languages (< %4) when culling low-frequency terms. From an informal interview with a native Chinese speaker, we believe this variance is derived from the diverse combinations of multiple basic color characters. For example, 蓝绿( ) is 蓝( ) plus 绿( ), 黄绿( ) is 黄( ) plus 绿( ). As for the orange hues 橘( ) and 橙( ), they are the names of different fruits – mandarins and oranges – and their use may reflect regional linguistic differences.

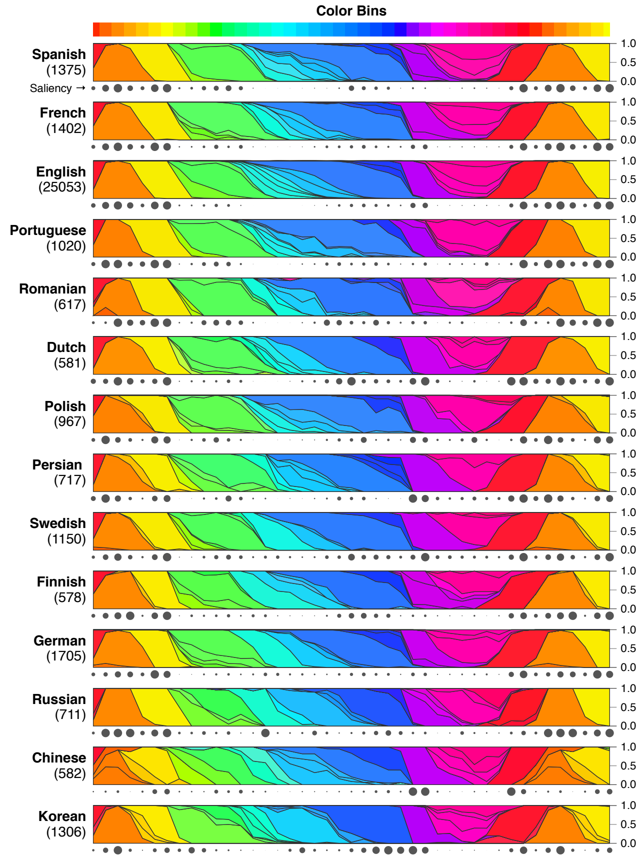

Figure 2: *The probabilities of terms for each hue color bin across 14 languages. Each area represents a term (t) and its height in a bin (c) represents P(t|c). The color of an area is the average color for the corresponding term. Gray circles below each chart encode each bin's saliency. Larger circles mean that the corresponding color is more likely to be called by a common name. The rows are sorted according to their distributional similarity.*

For the full colors in Korean and English, we observe the nameable color clusters in Figure 1. In English, we see 10 evident clusters that match Berlin & Kay's English basic color terms [BK69], as seen in prior name models [HS12]. Meanwhile, the Korean model exhibits additional clusters such as, 남 (near black), 청록 (between blue and green), 자주(between purple and red), 하늘 ( ), 연두 ( ), and 연보라 (between light blue and pink, a compound of the words 연[slight] and 보라[purple]). We also note that the models do not include a cluster for white, likely due to the white background (as discussed in previous work [HS12]).

4.2. Translation via Color Naming Model

We can use our models to quantify the translation quality of cross-language term pairs. We use the color-term probability $P(C|t)$ and assume that a translation should conserve this probability distribution over color bins. We formulate the translation loss for translating a term t_s in a source language l_s to another term t_t in a target language l_t as a distance:

$$translationLoss(t_s, l_s, t_t, l_t) = distance(P(C|t_s, l_s), P(C|t_t, l_t))$$

© 2019 The Author(s)
Eurographics Proceedings © 2019 The Eurographics Association.

We employ the Earth Mover's Distance [PW08, PW09] within CIELAB space as our distance metric. With this loss function, the best translation for a term t is expressed as:

$$translation(t, l_s, l_t) = \underset{t_t}{\operatorname{argmin}}\left(translationLoss(t_s, l_s, t_t, l_t)\right).$$

To evaluate this model, we compute translations between the top 100 most frequent terms in the Korean and English full color datasets. We then compare our translations to popular English-Korean online translation tools: Google Translate and Papago [LKS*16]. Figure 3 visualizes the translations of the major color terms (shown in bold) we previously identified in Figure 1.

The results reveal some questionable translations by the online translation tools: translations from purple ( ) to 자주 ( ) (and vice versa), from 연보라 ( ) to lightpurple ( ), from 남 ( ) to indigo ( ), and from 청록 ( ) to turquoise ( ) or cyan ( ) all have translation losses greater than the corresponding just noticeable distance (JND) in CIELAB space. In other words, these translated terms are predicted to have a $\geq$ 50% chance of being perceived differently from the source term.

5. Conclusion & Future Work

Our findings corroborate prior results about differences in nameable colors across languages, for example, that Russian has two salient blue colors [WWF*07]. With our data and models we can discover such color naming patterns across a number of languages. The differences we find can be used to further assess the generalizability of language-dependent differences in the perception of colors from previous studies [GPRMA17, Ath09, ADKS11, WWF*07].

The differences in nameable colors we observe across languages suggest extensions to prior color palette evaluations based solely on English color terms [HS12, GLS17]. Color palettes can be adapted specifically to speakers of different languages. For example, Korean and Russian have darker blue colors (파랑 and **синий**) as shown in Figure 2. A designer might use these darker blues to improve color saliency for Korean and Russian audiences. The improved saliency should hopefully promote those viewers' ability to reference and remember graphical elements.

In addition, our translation model predicts possible misunderstandings when verbally communicating about data visualizations across different languages. Consider a heat map that uses the *viridis* colormap [SvdW15]. If an English speaker refers to a "green" area, they may be referring to a wide range of colors in the colormap ( ). If a Korean speaker translates this as "초록" (following existing online translation tools), they may only consider a narrower range of colors ( ). To avoid this type of communication mismatch, visualization designers may wish to consider color name translation losses.

Looking ahead, there remains a range of improvements that could be made to our current study. The quality of the data might be further improved, as the cleaning of color terms was supervised by native speakers for only six of the languages. Collecting more non-English color-term pairs would permit fine-grained analysis for more languages, including translation models. Finally, larger datasets could support more nuanced models that also incorporate collected demographic data, for example to examine potential differences according to reported gender. We hope to investigate these possibilities as our color naming experiment continues to collect more data on LabintheWild.

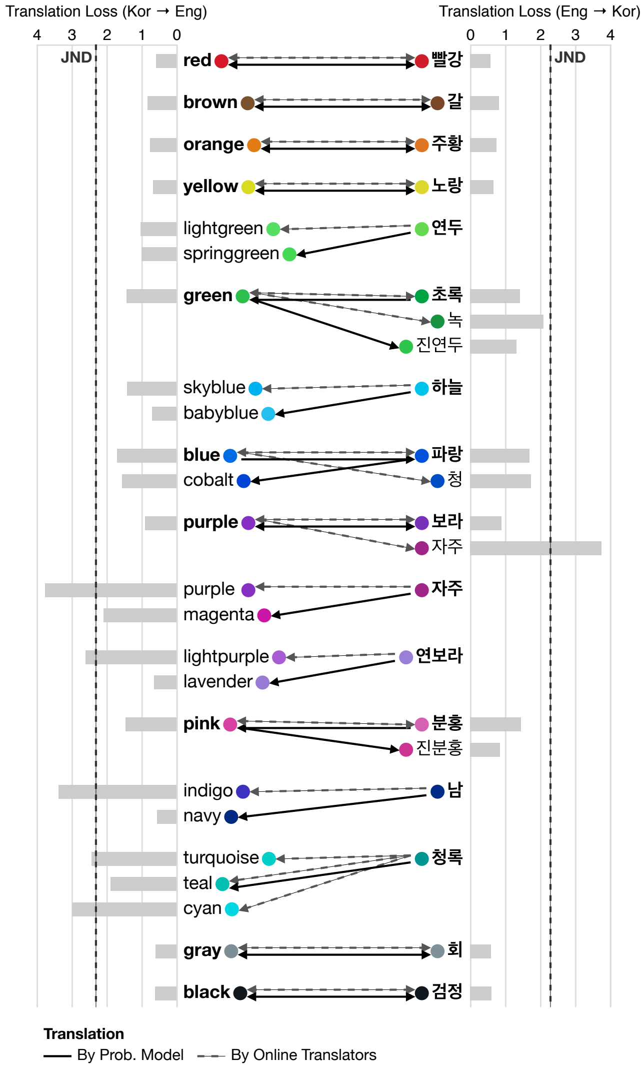

Figure 3: *Comparison of translations. Popular terms (bold) in English and Korean are translated by the probabilistic model (black arrows) and online translators (gray dashed arrows). The circles use the average color for the name, and the bars encode the translation losses. The vertical dashed line denotes approximately 1 just noticeable difference (JND) in CIELAB space [Sha02].*

6. Acknowledgements

Yvonne Chen and Jingjing Wang helped prototype the experiment and visualizations. Angli Liu, Tongshuang Wu, Nara Jung, Nigini Oliveira, Sayena Majlesein, and Yasaman Sefidgar provided translations of the experiment instructions and helped with data cleaning. We also appreciate feedback and help from Katharina Reinecke and the WildLab. This work was supported by a Moore Foundation Data-Driven Discovery Investigator Award.

© 2019 The Author(s)
Eurographics Proceedings © 2019 The Eurographics Association.

References

[ADKS11] ATHANASOPOULOS P., DAMJANOVIC L., KRAJCIOVA A., SASAKI M.: Representation of colour concepts in bilingual cognition: The case of japanese blues. *Bilingualism: Language and Cognition 14*, 1 (2011), 9–17. doi:10.1017/S1366728909990046. 1, 4

[Ath09] ATHANASOPOULOS P.: Cognitive representation of colour in bilinguals: The case of greek blues. *Bilingualism: Language and Cognition 12*, 1 (2009), 83–95. doi:10.1017/S136672890800388X. 1, 4

[BK69] BERLIN B., KAY P.: *Basic Color Terms: Their Universality and Evolution*. University of California Press, 1969. 1, 2, 3

[CSH08] CHUANG J., STONE M., HANRAHAN P.: A probabilistic model of the categorical association between colors. *Color and Imaging Conference 2008*, 1 (2008), 6–11. URL: https://www.ingentaconnect.com/content/ist/cic/2008/00002008/00000001/art00002. 1, 2, 3

[Far43] FARNSWORTH D.: The farnsworth-munsell 100-hue and dichotomous tests for color vision∗. *J. Opt. Soc. Am. 33*, 10 (Oct 1943), 568–578. URL: http://www.osapublishing.org/abstract.cfm?URI=josa-33-10-568, doi:10.1364/JOSA.33.000568. 2

[GLS17] GRAMAZIO C. C., LAIDLAW D. H., SCHLOSS K. B.: Colorgorical: Creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (Jan 2017), 521–530. doi:10.1109/TVCG.2016.2598918. 4

[GPRMA17] GONZÁLEZ-PERILLI F., REBOLLO I., MAICHE A., ARÉVALO A.: Blues in two different spanish-speaking populations. *Frontiers in Communication 2* (2017), 18. URL: https://www.frontiersin.org/article/10.3389/fcomm.2017.00018, doi:10.3389/fcomm.2017.00018. 1, 4

[HS12] HEER J., STONE M.: Color naming models for color selection, image editing and palette design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 1007–1016. URL: http://doi.acm.org/10.1145/2207676.2208547, doi:10.1145/2207676.2208547. 1, 2, 3, 4

[KBM*09] KAY P., BERLIN B., MAFFI L., MERRIFIELD W. R., COOK R.: *The World Color Survey*. CSLI Publications Stanford, 2009. 1, 2

[KR06] KAY P., REGIER T.: Language, thought and color: recent developments. *Trends in Cognitive Sciences 10*, 2 (2006), 51 – 54. URL: http://www.sciencedirect.com/science/article/pii/S1364661305003530, doi:https://doi.org/10.1016/j.tics.2005.12.007. 2

[LKS*16] LEE H.-G., KIM J.-S., SHIN J.-H., LEE J., QUAN Y.-X., JEONG Y.-S.: papago: A machine translation service with word sense disambiguation and currency conversion. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations* (2016), pp. 185–188. 4

[Mun10] MUNROE R.: Xkcd color survey results, 2010. URL: http://blog.xkcd.com/2010/05/03/color-survey-results/. 2

[PW08] PELE O., WERMAN M.: A linear time histogram metric for improved sift matching. In *Proceedings of the 10th European Conference on Computer Vision: Part III* (Berlin, Heidelberg, 2008), ECCV '08, Springer-Verlag, pp. 495–508. URL: https://doi.org/10.1007/978-3-540-88690-7_37, doi:10.1007/978-3-540-88690-7_37. 4

[PW09] PELE O., WERMAN M.: Fast and robust earth mover's distances. In *2009 IEEE 12th International Conference on Computer Vision* (Sep. 2009), pp. 460–467. doi:10.1109/ICCV.2009.5459199. 4

[RDD00] ROBERSON D., DAVIES I., DAVIDOFF J.: Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General 129*, 3 (2000), 369. 1

[RG15] REINECKE K., GAJOS K. Z.: Labinthewild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (New York, NY, USA, 2015), CSCW '15, ACM, pp. 1364–1378. URL: http://doi.acm.org/10.1145/2675133.2675246, doi:10.1145/2675133.2675246. 2

[Sha02] SHARMA G.: *Digital Color Imaging Handbook*. CRC press, 2002. 4

[SvdW15] SMITH N., VAN DER WALT S.: A better default colormap for matplotlib.(2015), 2015. URL: https://www.youtube.com/watch?v=xAoljeRJ3lU. 4

[WWF*07] WINAWER J., WITTHOFT N., FRANK M. C., WU L., WADE A. R., BORODITSKY L.: Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences 104*, 19 (2007), 7780–7785. URL: https://www.pnas.org/content/104/19/7780, doi:10.1073/pnas.0701644104. 1, 3, 4

© 2019 The Author(s)
Eurographics Proceedings © 2019 The Eurographics Association.