Title: Crowdsourcing of Multilingual Colour Names

Author

Dimitris Mylonas

Dept. Computer Science, University College London, United Kingdom

Synonyms

online experiments; collective intelligence; color language

Definition

Multilingual crowdsourcing refers to distributed experimental methodologies designed to collect large datasets of color naming responses in different languages from a large number of observers over the Internet.

Color naming

Color naming links vision and language. It describes the intriguing cognitive capacity of humans to communicate about regions of color space using category labels, for example, yellow, navy blue and dark olive green (MacDonald & Mylonas, 2016). A small number of basic color terms, corresponding to the English white, black, red, yellow, green, blue, brown, orange, pink, purple and grey, are shared and comprehended well by most speakers in many languages (Berlin & Kay, 1969/1991). Yet, color names vary across languages, lexically, in number and in range of reference. To augment color communication within and between different languages, it is necessary to have a worldwide method for gathering the words and phrases that people use to describe a wide range of colors (Cook, Kay & Regier, 2005; Mylonas & MacDonald, 2010).

Color systems are usually three-dimensional geometric spaces that allow the description of colors with numerical coordinates. The representation of perceived colors in such spaces can be very useful for specifying the referents of color names in terms of their coordinates. Using color systems in cross-cultural research enables measurements of differences and similarities of color categories across languages. In Figure 1 are shown subdivisions of a continuous color plane into English and German lexical color categories, based on the responses of human observers in a color naming experiment (Mylonas, Stutters, Doval & MacDonald, 2013). It is interesting that not only the names are different (language-dependent) but also the boundaries of the corresponding regions. For example, the region for *hellblau* in German is larger than the region for *light blue* in English.

Figure 1. Classification synthetic image (left) by observer-based color names in English (middle) and German (right). Each category is represented by the color of its centroid (Mylonas et al., 2013).

Crowdsourcing and laboratory color naming studies

Human-centered color naming studies are traditionally performed in controlled laboratory settings and involve predefined naming lists (e.g. restricting the responses to a limited number of single word terms) and labor-intensive tasks for a small number of observers (Boynton & Olson, 1987; Sturges & Whitfield, 1995; Mylonas, Griffin & Stockman, 2019). Despite the usefulness of calibrated experiments in the field of color science, experimenters face several limitations. First, the controlled viewing conditions, while advantageous for accurate colorimetric specification of stimuli, limit the validity of the predicted color naming functions in real-world settings. Second, the pool of available observers is often constrained to a small number of college students or to the authors of the study, which makes it difficult to generalize the results to heterogenous groups of the population. Third, the large individual differences reported in color naming studies makes the generalization of laboratory experimental findings open to doubt. Also, constrained color naming methods are able to capture only a small fraction of the richness of color languages of the world and the labeling of large regions of color space remains ambiguous. For example, Figure 2 shows the percentage of number of words used by subjects in an unconstrained color naming experiment (Mylonas & MacDonald, 2010). Constraining the responses only to the 11 basic color terms would account for 29% of the responses while to single words would account only for 52% of the responses for naming stimuli across the full color gamut.

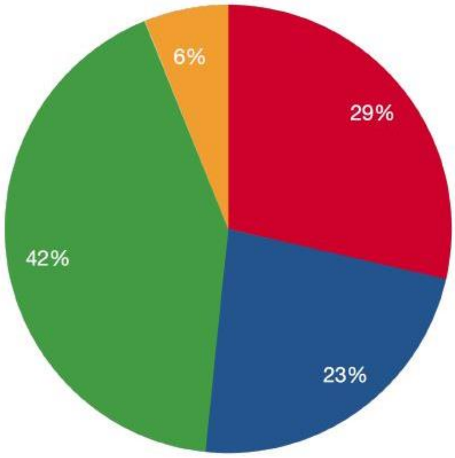

Figure 2. Number of words used by subjects in an unconstrained color naming experiment in English. In red proportion of 11 basic color terms, in blue single words non-basic terms, in green two-word color names and in orange color names consisting of three-words or more (Mylonas & MacDonald, 2010).

On the other hand, web-based color naming experiments provide greater ecological validity than traditional approaches by allowing simultaneous participation of observers in their own familiar spaces, in their own time, with their own equipment and without the physical attendance of the examiner (Moroney, 2003, Mylonas & MacDonald, 2010). They also avoid the restrictions of the usual methods which have a small number of observers and/or a limited set of monolexemic color terms. Instead, thousands of observers from linguistically and demographically diverse populations are able to name freely a large number of colors online and produce larger color lexicons that improve the precision of color names. Online methods also depart from previous research by distributing the color naming task over a large population. In the collection of multilingual color naming data, online experiments extend earlier cross-cultural studies, which typically used only the most saturated color samples on the surface of the Munsell system, by also sampling the interior of the color solid (Berlin & Kay, 1969/1991; Cook et al., 2005). Yet, online experimental methodologies often receive criticism as not meeting the exacting standards demanded for rigorous color research because of the uncalibrated color reproduction and viewing conditions. For example, different participants may use different display technologies and browsers and they may view the color stimuli from different distances and viewing angles under various illumination conditions (Moroney, 2003). Furthermore, the motivation of participants varies in both reward-based and in volunteering-based crowdsourcing, and screening the quality of the responses is a major challenge. However, the practical success of sRGB in color encoding for Internet applications has greatly improved the consistency of displayed colors and people exhibit a remarkable ability to describe consistently the color of objects in a continually changing visual world. The robustness of crowdsourcing experimental results can be assessed by whether or not their variability is larger than the variability in laboratory-based studies (Moroney, 2003; Mylonas & MacDonald, 2010; Mylonas, Griffin & Stockman, 2019).

Crowdsourcing color naming experiments

To understand the workings of crowdsourcing a number of online color naming experiments are briefly reviewed. Nathan Moroney (2003) was the first to publish the results of an unconstrained web-based color naming experiment. More than 700 participants were asked to give the best names for 7 patches of color selected randomly from a 6 × 6 × 6 uniform grid sampling of the RGB cube, viewed on their own displays against a white background. Figure 3 shows a simplified flowchart of this experiment to create a dynamic website that presents colors and collects color naming responses.

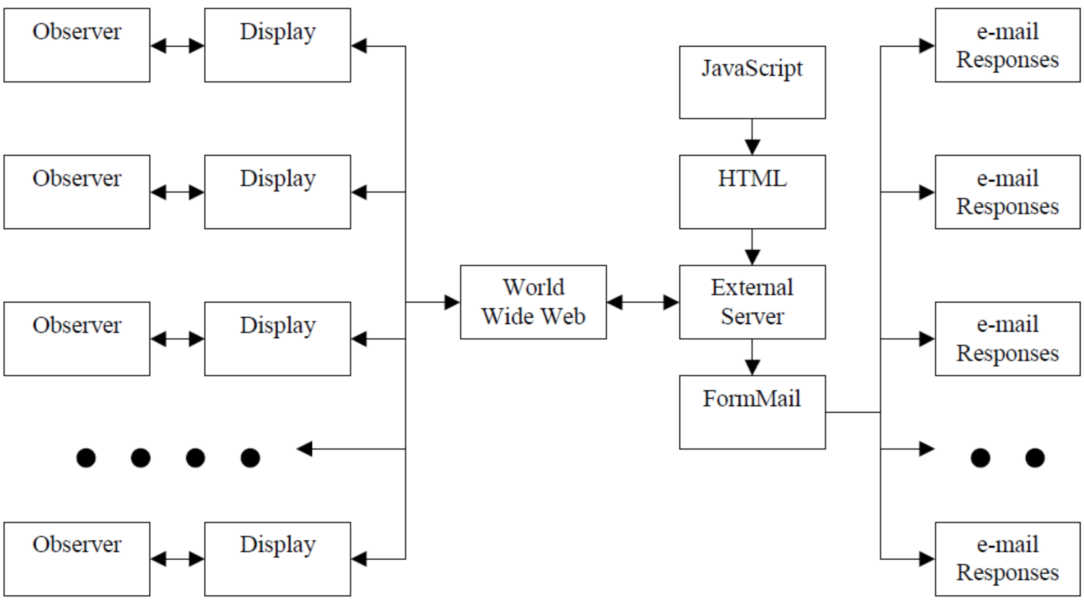

Figure 3. Overall flowchart of web-based color naming experiment (Moroney, 2003).

Findings of the online experiment of Moroney (2003) were validated against the results of Boynton & Olson (1987) and Sturges & Whitfield (1995), both obtained under controlled laboratory conditions. The comparisons showed a high degree of correlation for the chromatic basic color terms in terms of hue angles in CIELAB with a linear fit of $R^2 = 0.99$ and $R^2 = 0.98$ respectively. The same comparison but for 27 common color names against the results of a subsequent web-based study resulted in a linear fit of $R^2 = 0.99$ (Mylonas & MacDonald, 2010). In 2008, DoloresLabs publicized the results of an online color naming survey conducted via Mechanical Turk to collect a small number of responses for a large number of colors. In the raw dataset, a total of 1966 distinct color descriptions were collected for 10,000 randomly sampled colors in the RGB cube 9. An analysis of the DoloresLabs dataset by Chuang and his colleagues (2008) showed a good correspondence for colors on the surface of the Munsell system with the results of the World Color Survey (WCS) for red, brown and purple but hue shifts for blue and green (Cook et al., 2005). The analysis of data across the full color gamut revealed two additional consistent categories at the lavender and cyan regions that do not correspond to any salient region in the WCS data (Figure 4).

In 2010, the author of the web comic XKCD, Randall Munroe collected a dataset of 3.4 million unconstrained responses mainly in English. This very large dataset is associated with metadata regarding the gender, language skills, color-blindness of the subjects, and information about their monitor settings. Observers were free to name as many sets of color swatches as they liked, presented against a white background. Each color swatch was uniformly sampled from the full RGB cube. An analysis of the publicly available dataset by Heer & Stone (2012) showed clusters that correspond well to the basic color terms identified by Berlin & Kay (1969/1991).

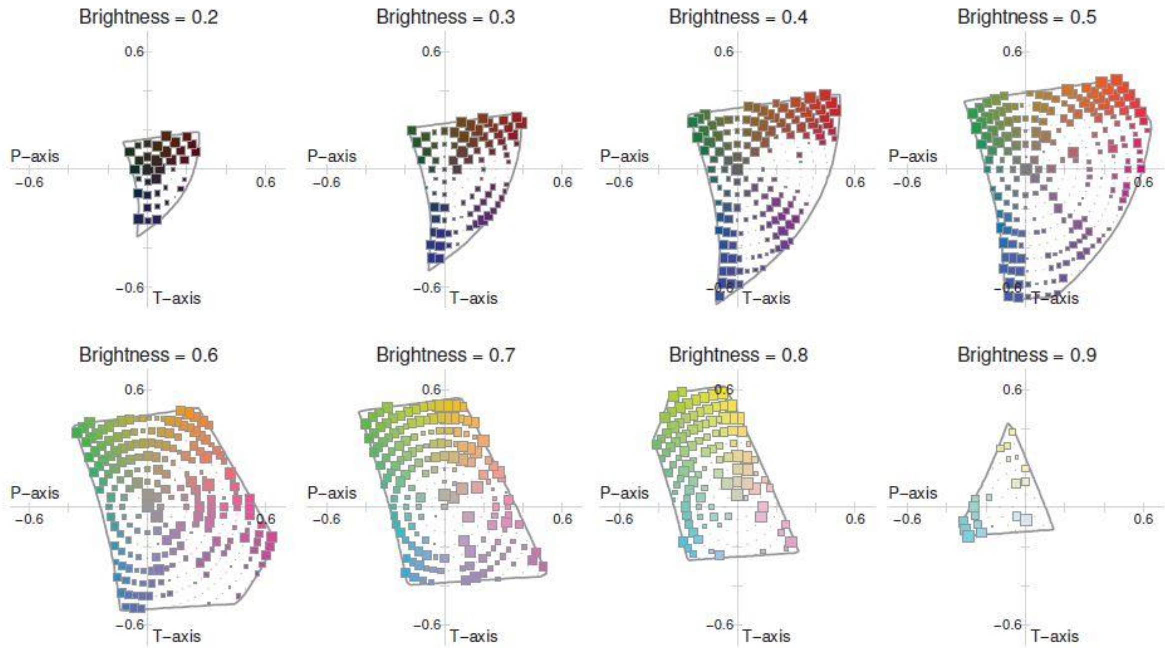

Figure 4. Consistency of color naming responses on 8 brightness levels in IPT space. The size of each colored square corresponds to the degree of agreement between observers for the name of this chip (Chuang et al., 2008).

In 2009, Mylonas & MacDonald launched an ongoing color naming experiment to collect broad sets of color names within different languages with their corresponding color ranges in sRGB and Munsell specifications over the Internet (Mylonas & MacDonald, 2010). Each observer was presented with a sequence of 20 color patches randomly selected from 600 simulated Munsell samples, presented against a neutral grey background. The color naming responses were associated with metadata regarding the response time, color deficiency, hardware/software components, viewing conditions, gender and cultural background of the observers. An analysis of the data for the most important color names suggested the extension of the English inventory from the 11 basic color terms of Berlin & Kay (1969/1991) to 13 basic terms including a lilac and a turquoise category (Mylonas & MacDoland, 2016; see also Chuang et al., 2008). Except blue, comparisons of centroid location of the basic color terms in different languages showed a good correspondence (>5ΔE_{00}) despite the online experimental methodology and the linguistic diversity of the observers (Mylonas et al., 2013; Paramei, Griber & Mylonas, 2018). However, the agreement between British and American English (ΔE_{00} = 1) was better than the agreement between English and non-English color languages (Mylonas, MacDonald & Griffin, 2017). Figure 5 shows the location of the thirty most frequent color names offered by a large number of British and American English speakers in the online color naming experiment. The online experiment was also validated directly against an unconstrained color naming experiment using the same stimuli on a calibrated CRT monitor in controlled viewing conditions (Mylonas, Griffin & Stockman, 2019). The agreement (<5ΔE_{00}) for the location of the basic color terms between the web- and laboratory-based experiments was satisfactory, and superior to the agreement (>7ΔE_{00}) between previous laboratory-based studies (Boynton & Olson, 1987; Sturges & Whitfield, 1995).

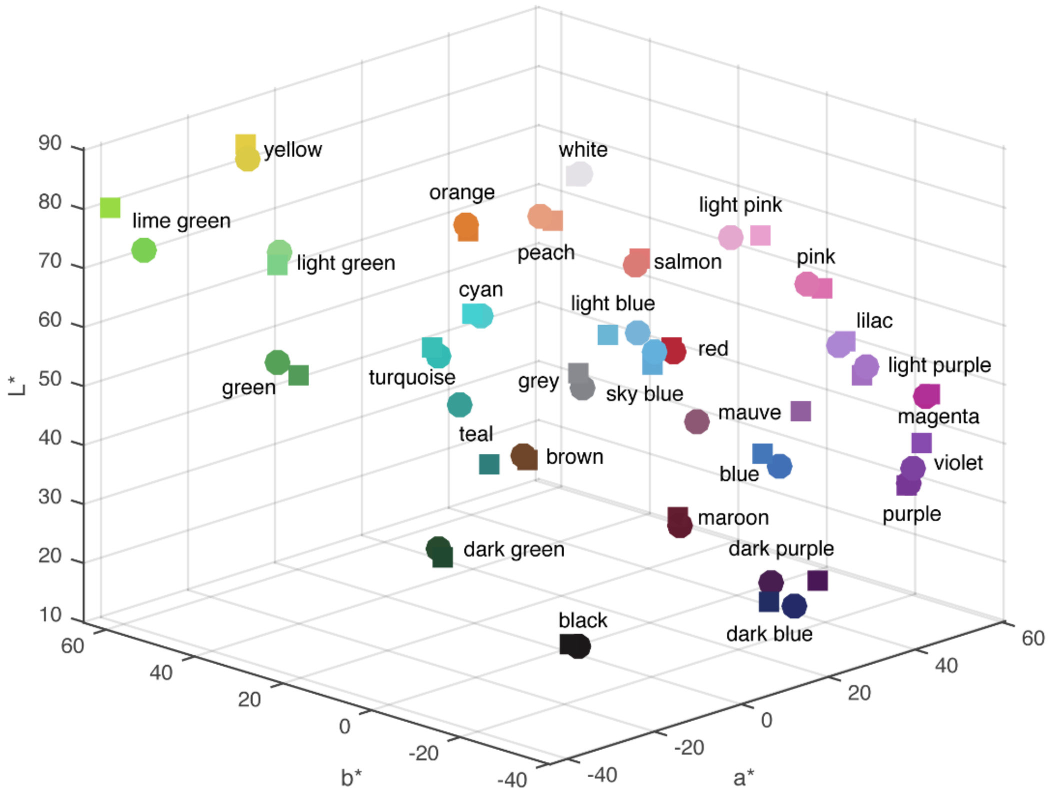

Figure 5. Location of centroids of 30 most frequent color names in CIELAB for British (square) and American (circle) English speakers (Mylonas et al., 2017).

Conclusions

Crowdsourcing experimental methodologies offer considerable advantages over traditional approaches for obtaining richer color naming datasets in multiple languages. Comparisons between the results of web- and laboratory- based studies show that the locations of color categories correspond well and support the validity of both methods in estimating color naming functions in calibrated and uncalibrated monitor settings. Moreover, the overall good agreement between the location of corresponding basic color terms in multiple languages confirms that speakers of different languages tend to categorise colors into basic color terms similarly but differences do exist, especially in the blue region and between non-basic categories. New mobile display technologies and online platforms offer new capabilities and present new challenges to researchers for conducting color naming research over the Internet.

Cross-References

Color Lexicons; Cross Cultural Communication; Psychological color space and color terms, Berlin & Kay Theory

References

1. Berlin, Brent., & Kay, P. *Basic color terms: Their universality and evolution*. Stanford, Calif.: Center for the Study of Language and Information (1969/1991)
2. Boynton, R. M., & Olson, C. X. Locating basic colors in the OSA space. *Color Research & Application*, *12*(2), 94–105 (1987)
3. Chuang, J., Hanrahan, P., & Stone, M. *A Probabilistic Model of the Categorical Association Between Colors*. 6–11. Portland, Oregon, USA: The Society for Imaging Science and Technology (2008)
4. Cook, R. S., Kay, P., & Regier, T. The World Color Survey Database. In H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science*, 223–241 (2005)
5. Heer, J., & Stone, M. Color naming models for color selection, image editing and palette design. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems*, 1007–1016 (2012)
6. MacDonald, L., & Mylonas, D. *Colour Naming – Linking Vision and Speech*. 16–25. Krakow, Poland (2018)
7. Moroney, N. Unconstrained web-based color naming experiment. *Color Imaging VIII: Processing, Hardcopy, and Applications. Edited by Eschbach, Reiner; Marcu, Gabriel G. Proceedings of the SPIE*, *5008*, 36–46 (2003)
8. Mylonas, D., & MacDonald, L. Online Colour Naming Experiment Using Munsell Samples. In *5th European Conference on Colour in Graphics, Imaging, and Vision and*

12th International Symposium on Multispectral Colour Science. CGIV 2010/MCS'10, Joensuu, Finland (2010)

9. Mylonas, D., Stutters, J., Doval, V., & MacDonald, L. Colournamer, a synthetic observer for colour communication. In L. MacDonald, S. Westland, & S. Wuerger (Eds.), *Proc. 12th Congress of the International Colour Association*, 701–704 (2013)
10. Mylonas, D., & MacDonald, L. Augmenting basic colour terms in English. *Color Research & Application*, *41*(1), 32–42 (2016)
11. Mylonas, D., MacDonald, L., & Griffin, L. D. Differences in Color Naming between British and American English Speakers. *Proc. 13th AIC Congress*. Jeju, South Korea (2017)
12. Mylonas, D., Griffin, D. L., & Stockman, A. *Mapping colour names in cone excitation space*. In 25th Symposium of the International Color Vision Society, Riga, Latvia (2019)
13. Paramei, G. V., Griber, Y. A., & Mylonas, D. An online color naming experiment in Russian using Munsell color samples. *Color Research & Application*, *43*(3), 358–374 (2018)
14. Sturges, J., & Whitfield, A. Locating basic colours in the Munsell Space. *Color Research and Application*, *Volume 20* (6), 364–376 (1995)