

E

Effects of Color Terms on Color Perception and Cognition



Jonathan Winawer¹ and Nathan Witthoft^{1,2}

¹Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

²Department of Psychology, Stanford University, Stanford, CA, USA

Synonyms

[Categorical perception](#); [Linguistic influences on color perception](#); [Sapir-Whorf hypothesis](#)

Definition

The possibility that naming colors, either in a single instance or habitually over a lifetime, alters color perception.

Color Perception and Color Communication

When we communicate about the colors of scenes and objects comprising our visual experience, what we see informs our choice of words. A question that has interested many cognitive psychologists is whether the color words we use

affect how we see. One view is that naming is for communication and has no effect on how we see or experience the world. In this view, color appearance, and therefore performance on tasks which strictly depend on color appearance, is determined entirely by the visual system and not at all by the language that one speaks. Color vision is informationally encapsulated, its output is automatically produced, and while its output is available for cognitive processes like decision making and speaking, higher cognitive processes cannot alter its process [1]. This can be contrasted with the idea that the very act of naming a color, either in a single instance or habitually over a lifetime, can change one's perceptual experience with the result that colors assigned the same label (i.e., belonging to the same color category) are more difficult to discriminate than colors assigned different labels (belonging to different color categories).

The two views have persisted in the psychology literature for at least two reasons. First, the empirical evidence has been mixed, with some researchers finding effects of color terms on perceptual or cognitive tasks and others failing to find such effects. Second, when positive evidence of the influence of color naming on color perception has been found, it has often been unclear at what level the effects are manifested. For example, the same experimental results might be interpreted by some researchers as evidence that color terms affect how one sees colors and by other researchers as evidence that color terms affect

how one remembers or makes judgments about colors. In this entry, we examine the question of how to design experiments to probe the effects of color terms on color perception and cognition and how to interpret the effects.

The study of categorical effects in color perception has its roots in two historical traditions. The first is the Sapir-Whorf Hypothesis, a concept that originated in anthropology [2]. Whorf and Sapir formulated this hypothesis in different ways in their writing, but researchers have generally taken their view to be something like the way in which experiences are mapped onto words affects the experiences themselves, so that speakers of different languages perceive and conceive of the world differently. A second and related historical tradition is the study of categorical perception [3, 4] and in particular the study of phoneme perception in language. Distinctions between certain phonemes, such as /r/ and /l/, are necessary for speaking some languages, such as English, but not other languages, such as Japanese. Adult English speakers are better at the /r/ versus /l/ discrimination than Japanese speakers (though Japanese speakers are above chance, meaning that the effects are not completely categorical). In phoneme learning, it appears that the learning is primarily a loss rather than a gain; the /r/ and /l/ sounds become more similar as a child learns Japanese rather than more different as a child learns English [5]. Psychologists studying color have similarly asked whether making color distinctions improves sensitivity to distinguish shades that straddle a category border or worsens one's sensitivity to distinguish shades that fall within a single color category.

No color scientist takes seriously the idea that color perception is as categorical as color naming. It is trivial to find two colors that are both called green that can be distinguished from one another. A more nuanced question is whether colors that are named alike show any degree at all of increased perceptual similarity (or decreased discriminability) compared to colors that are named differently.

Breaking the Circularity: The Logic of Measuring Category Effects

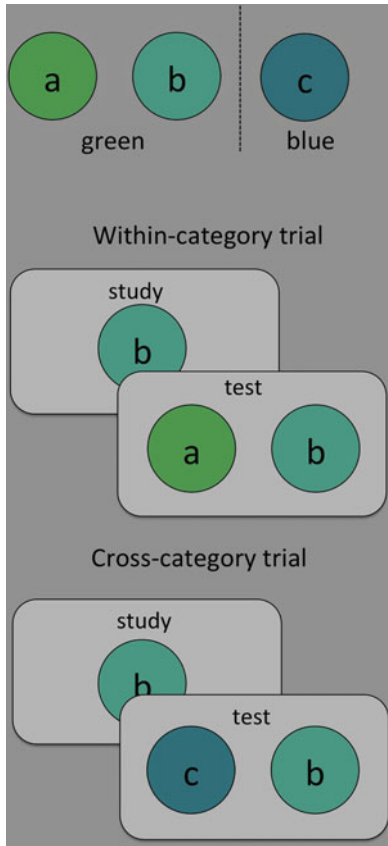
To assess the effects of color categories on perception and cognition, one has to sort out the direction of causation: a pair of colors may belong to different categories because they look very different, rather than looking different because one learns to put them in different categories. In fact, it is certain that this happens; colors that look alike are more likely to belong to the same color category, and language groups across the world tend to assign colors to categories in similar ways, though some differences exist [6]. To know whether there is *also* an effect in the reverse direction, that is, an effect of category on perception or cognition, one needs to have a prediction of how similar the colors should be in the absence of a category effect. Given such a prediction, one could ask the following: are two colors from the same category (Fig. 1a, b) less discriminable (or more similar) than two colors from different categories (Fig. 1b, c), even when the two pairs are identically spaced? That is, are “within-category” trials harder than “between-category” trials? This leads to a potentially circular problem. If the stimulus pairs are chosen such that they are equally spaced, then are they not guaranteed to be equally similar? Or, conversely, if one pair is shown to be more discriminable than the other, does this mean that the color space which placed them at equal distances from one another was flawed?

This entry summarizes the answer to two questions pertaining to whether and how color naming affects color cognition:

1. How are experiments designed to break the problem of circularity posed above?
2. If the problem of circularity can be broken, then what aspects of color perception and cognition show effects of color names?

Model-Based Approaches

One way to break the circularity problem in studying the effect of color names on color cognition and perception is to compare the results to a null model, meaning a model which does not use color



Effects of Color Terms on Color Perception and Cognition, Fig. 1 Testing effects of color terms on a color task. The *upper* panel shows three stimuli that might be used in an experiment to investigate whether color terms affect a particular task. If the task involves memory, the subject may be instructed to first remember a color (“study”) and then identify the remembered color (“test”). If the incorrect stimulus comes from the same category, the trial is considered “within-category.” If the incorrect stimulus comes from a different category, it is considered a “cross-category” trial. The colors were chosen so that *a* and *b* look *green* and *c* looks *blue*, but the observer may not agree depending on the display and individual differences between observers

categories as part of its machinery even as it tries to predict them. The choice of color space for a null model is an essential consideration. If the space is designed to reflect perceptual similarity, such as the Munsell color space, then it may already incorporate category effects. And if such a space fails to predict performance on a color judgment task, then it may simply indicate that the space did not perfectly achieve its design goal,

rather than something about color categories. Hence, a null model is of limited value for the study of effects of language on color processing if the model is purposely shaped to account for the pattern of human behavior, such as the Munsell system or the CIE Lab space.

In contrast, a model that sticks more closely to peripheral aspects of color processing can be informative. For example, Brown and colleagues used the MacLeod-Boynton space to predict reaction time in a visual search experiment for an oddball color [7]. The MacLeod-Boynton space represents color stimuli as a linear transform of cone excitations. It contains no information about color names or categories or other properties of color cognition. In the visual search experiments, one color (the target) differed from the other colors (the distractors), and the time to find the target was modeled as a function of the difference between the target and distractors in MacLeod-Boynton space. Trials in which there was a large difference between the targets and distractors in the space were predicted to have a faster reaction time. Some trials were cross-category and some within-category, analogous to the design in Fig. 1. The authors found that the null model provided a good fit to the empirical results, with no systematic deviations for cross-category trials. For this color task, color discrimination time could be explained by a model based only on cone absorptions, and hence there was no evidence that discrimination was affected by color categories. The null model succeeded.

It is important to consider what happens when a null model fails. Suppose, for example, a model predicts that stimuli *a* and *b* and stimuli *b* and *c* in Fig. 1 are equally discriminable, and yet experiments show that cross-category pairs are discriminated faster or recalled more accurately than within-category pairs. While additional mechanisms or a different model may fit the data better, it does not follow that color categories made the stimuli more discriminable. It may simply be that the model was incorrect or inappropriate for the task.

Equating Color Stimuli but Varying the Subject Population

Another way to break the circularity problem in studying the effect of color names on color cognition and perception is to test identical sets of color stimuli with speakers of languages with differing color terms.

For example, for English speakers there is no common color word that refers to all three stimuli in Fig. 1. Speakers of some languages (sometimes called “grue” languages), however, could refer to all three colors by the same name. An experimenter can therefore avoid the difficulty of perfectly equating differences between the stimulus pairs: for speakers of languages like English that place a boundary between stimuli b and c, the b/c pair should be more different than the a/b pair *relative to* speakers of “grue” languages. The colors are controlled not by the exact localization of colors in a color space but rather by showing exactly the same sets of colors to different groups of observers and seeing if their pattern of responses differs. One of the challenges in such experiments is finding appropriate subject populations. Color naming patterns are quite similar across many languages. Nonetheless, large differences in naming patterns can be found by comparing populations from an industrialized western society such as the UK or the USA with speakers from more remote cultures which have had minimal contact with western societies, such as the Dani and Berinmo of Papua New Guinea or the Himba of Namibia. However, even among modern industrial societies, differences can exist. For example, Russian speakers use different words for the colors light blue (“goluboy”) and dark blue (“sinij”), while English speakers may refer to both as “blue.”

The results of cross-linguistic studies have not been uniform. Some have found effects of language on a color task, such as a color memory task comparing English speakers with Berinmo speakers [8]. Others have found little to no effect. One reason is that the effects, even when found, tend to be small compared to the shared properties of color processing found in all observers. For example, consider the effect of metamers: two colors with quite different spectral power

distributions will be indistinguishable to an observer if the two colors result in the same pattern of cone excitations. This is a very large effect and is quite consistent across observers. A metameric pair for one person is likely to be a metameric pair for another person, or nearly so, unless one of the observers has a color vision deficit. Put another way, three-channel color displays such as television do not need to be tailor-made according to the color naming patterns within a language community. Color metamers do not differ with language because the information lost at the level of cone absorptions cannot be regained. The cross-linguistic differences that have been observed tend to be quite modest in magnitude relative to the much larger effects shared across all observers.

Equating Color Stimuli but Varying the Hemifield: Lateralized Whorf

Rather than testing for an effect of language by holding the stimuli constant and varying the language of the subjects, some experimenters have compared the same speakers but on different sides of visual space. In these experiments, ostensibly the same set of color stimuli are shown on either the left or the right side of a fixation point. The logic of this kind of experiment is that verbal processes are computed primarily in the brain’s left hemisphere, and visual inputs are routed to the contralateral hemisphere, so that stimuli to the left of fixation are processed more in the right hemisphere, and vice versa. The separation by hemisphere is not complete for either vision or language; nevertheless, researchers hypothesized that if effects of color terms on color tasks were to be found, they would be larger for stimuli presented to the right of fixation than the left of fixation. A number of reports have supported this hypothesis, known as the lateralized Whorf effect [9]. Other reports have failed to confirm the finding, however, and the reasons for the discrepancies have not yet been resolved [7, 10].

Equating Color Stimuli but Varying the Role of Language: Verbal Interference

Another strategy to break the circularity of measuring a category effect is to use the same set of

stimuli under different conditions designed to make it easier or harder to use language in the task. For example, a color task can be conducted under conditions in which the subjects' verbal faculties are engaged with a secondary task, such as reading a list of words aloud ("verbal interference"). If a secondary, language-related task reduces a putative category effect, then the problem of circularity is solved for this experiment. There is no need to precisely equate the distances between color pairs; rather, the role of language is implicated because performance changes when the availability of language changes. Experiments have shown that manipulating the availability of language during a color task can reduce the effect of color terms on the task.

The classic experiment of this type was done by Kay and Kempton [11], who compared English speakers to Tarahumara speakers from northern Mexico. The Tarahumara use the word "siyname" to name both green and blue. Subjects from both groups were shown triads of color chips which straddled the blue/green boundary and asked to select the chip which was least similar to the other two. They found that English speakers tended to exaggerate the distance between chips which straddled the boundary, while the Tarahumara did not. This finding is consistent with the idea that the availability of the terms "blue" and "green" influenced the English speakers' decisions about which color chip was the most dissimilar when one of the chips came from a different category. Kay and Kempton proposed that rather than resulting from some difference in color appearance, the English speakers were relying on the linguistic distinction to make decisions when the perceptual information was insufficient. To test this idea, they repeated the task, but only allowed the subjects to compare two of the chips at a time. The pairs were set up so that the chip in the middle of the color range could be alternately compared with one of the two to either side. This meant that if the middle chip was at a boundary, it would seem bluer than one chip and greener than the other, thus eliminating the usefulness of having a verbal code. Under these conditions, the English speakers showed the same performance as the Tarahumara.

If a category effect goes away when labels become unavailable or not useful, then it is unlikely that the effect is due to color terms affecting early perceptual processes. While such an account is logically possible, it would require color appearance to be altered only during those moments when one is accessing the labels. A more parsimonious explanation is that the *decision* process is affected by language. Verbal labels may be used to help keep track of the various stimuli in an experiment, either over a memory delay or when comparing stimuli spread over space. If, on a particular trial, all the stimuli come from the same verbal category (e.g., they are all blue), then labels are unlikely to help accomplish the task (and might even hinder performance). In contrast, if stimuli in a trial can easily be assigned with different labels (e.g., one blue and one green), then access to the labels may facilitate memory or the comparison process. If a verbal dual task interferes with the ability to label stimuli, even implicitly, then this may eliminate one strategy or source of information for accomplishing the task and hence may change performance. Thus, verbal interference effects are more likely to reflect a role of color terms on decisions, strategy, and memory, rather than perception.

Cognition and Perception: What to Measure

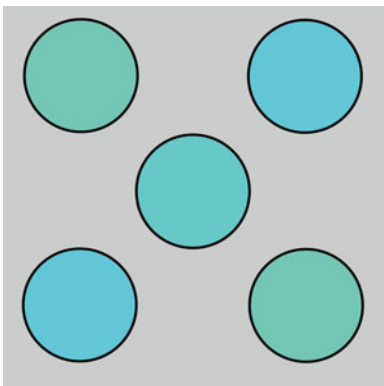
Thus far we have primarily been focused on how to break the problem of circularity when measuring the effect of color terms on color tasks. Here we focus on the tasks themselves. What kinds of tasks are used to test for the effects of color terms, and what do the results show?

Subjective Judgments

Perhaps the most natural way to ask whether color terms affect color perception is to ask subjects to make direct, overt judgments about color stimuli. For example, one could present subjects with a variety of color samples and ask them to place into groups those that are most similar. Experiments like these have been conducted with observers

across languages. Grouping patterns tend to be similar across language groups, but some differences are found consistent with grouping together colors that are given the same name in the language. The difficulty with this type of result is that subjects may choose to group together colors because they have the same name, rather than because they look most alike. Hence, it is difficult to identify whether the effect of color terms is on perception, or on the strategy of how to form groups.

Perceptual Grouping. A more indirect approach can make it less likely that subjects use a naming strategy to make a decision. For example, Webster and Kay took advantage of the tendency of human observers to perceptually group together objects that have a similar color [12]. By arranging disks into two diagonals with an ambiguous central disk, participants could indirectly report perceptual similarity by indicating whether the disks grouped into a line sloping up to the right or up to the left (Fig. 2). The question they then asked was whether the decision to group dots into a triplet was influenced by color categories, specifically whether there was a bias to group the dot with the diagonals when the three came from the same color category, regardless of the proximity in color space. These effects were compared to a null model based on a linear transform of cone excitations, similar to the approach used by Brown and colleagues [7]. The null model did a good job explaining the data, and hence there



Effects of Color Terms on Color Perception and Cognition, Fig. 2 Gestalt grouping configuration used to study the effect of color categories on color appearance [12]

was no evidence that color categories modulated the task.

Hue Scaling. Interestingly, a different measure of color perception using the same null model did show a significant deviation from the model. This measure is called “hue scaling,” in which subjects view a color sample and rate the perceived amount of red, green, blue, and yellow in a stimulus. Compared to a null model, colors that fell on the blue side of the blue/green border were effectively rated as containing more blue than would be expected purely from the model. That is, the results were biased away from the null model in the direction predicted by an effect of color terms on the task.

Subjective judgments are a desirable way to measure perception because they are direct: if one wants to know what something looks like to a subject, then asking for a subjective judgment is a sensible way to find out. On the other hand, because the judgments are subjective (there is no right answer) and there is no time pressure, these tasks are also amenable to strategic decisions that have little to do with perception. For example, when subjects are explicitly asked to rate the amount of blue in a stimulus, the internal mapping from perception to a judgment may be influenced by the fact that the subject knows that the stimulus falls in the category “blue.”

Memory

Memory tasks designed to probe the effect of color terms on perception or cognition often take the form of remembering a color over a delay and then trying to choose between the same stimulus and a second stimulus, where on some trials the second stimulus comes from the same category and sometimes a different category. Memory tasks have sometimes been used in cross-linguistic studies and sometimes in studies that included dual tasks (verbal interference). Several memory studies have found effects of color category on performance (e.g., [8, 13]). These studies are consistent with the idea that color terms are used as a dual code in memory; if the subject remembers the term in addition to the appearance, then the subject will be more accurate when the memory test includes only one option from the expected color category.

Reaction Time on Suprathreshold Discrimination

In a suprathreshold discrimination task, subjects typically judge which color patch matches a test color. There is only one right answer, and the discrimination is usually well above threshold so that subjects are expected to make the right answer most of the time. The measurement of interest is reaction time, and the typical experimental question is whether reaction time is faster for cross-category trials than within-category trials. In order to avoid the circularity problem, the experimenter can use one of the strategies described above, such as cross-linguistic comparisons, dual tasks, or varying the visual hemifield to which the stimuli are presented. Alternatively or in combination with these strategies, the experimenter can provide a null model to predict the data.

Suprathreshold experiments have had mixed results. Some have reported an effect of color categories on performance and some have not. When an effect of color categories is found in a speeded suprathreshold discrimination task, there are typically two alternative explanations the experimenter is confronted with. The participant may be faster on cross-category trials because stimuli that come from different categories look more different or the participant may be faster on cross-category trials because a decision is easier to make when color labels support the decision. As discussed above, if a verbal dual task eliminates a category effect, then the effect probably is in the decision stage rather than in the percept.

Subjects' experience with the task may also modulate the contribution of color categories on performance. For example, Witzel and Gegenfurtner [14] found that suprathreshold judgments were facilitated by category differences between stimuli at many name boundaries. These effects were not apparent in a second group of subjects tested on the same stimuli, but who were trained ahead of time on difficult boundary discriminations. The trained group was also faster and more accurate on the suprathreshold task. One interpretation is that categorical labels were not useful in the training phase, and therefore these subjects learned to make the

decision without waiting for the support of a categorical label.

Discrimination Thresholds

Threshold discrimination experiments are among the least ambiguous experiments in psychology. If an observer can discriminate two stimuli, then we can be certain that the observer's perceptual system has encoded the two stimuli differently. If the stimuli are indistinguishable (below threshold), then information distinguishing the stimuli was either not encoded or was lost in subsequent processing. If discrimination thresholds were altered by the color terms in one's language, this would provide the most direct evidence that color terms affect perception of colors. Experiments that have compared threshold discrimination among speakers of languages that divide the color spectrum differently have found no effect of color terms [15].

Interpretation and Future Directions

We began with the question of whether the words we use to describe colors affect how we see. A major challenge in addressing this question is sorting out the direction of causality: since similar colors are likely to be given the same name, and dissimilar colors are likely to be given different names, how does one measure whether naming has any effect on how the colors appear? Several methods have been used to address this problem, including the use of null models, cross-linguistic experiments, verbal interference, and hemifield manipulations.

Despite some discrepancies in the literature, some general patterns are emerging. A variety of tasks and experiments have reported effects of color terms on color cognition in the domains of judgment and decision making and memory. Results are consistent with the idea that color terms can serve as a dual code in conjunction with sensory representations for certain kinds of judgments. Consider the case of speeded discrimination tasks, discussed earlier. In these tasks, there are two sources of information that can inform a decision whether two colors are the

same or different. One is a sensory representation and the other is a label. In some situations, for example, when reaction times are slow and the two samples belong to different categories, the labels are useful and may speed reaction times. In other situations, such as two samples having the same color label, or subjects who are trained to respond faster than the labels arise, the categories are not useful and therefore will not contribute to the decision or may even interfere. This kind of explanation – that a dual code explains many effects of color terms on task performance – implies that the effect of color terms is on the perceptual decision-making process, and not on the sensory representations.

In contrast, there is no firm evidence that color appearance or color threshold discrimination is affected by color terms. The pattern of results in this way differs from phoneme discrimination. This may reflect the fact that phonemes have a categorical function: words are combinations of phonemes, and if a phoneme is mis-categorized, then the word may be misheard. In contrast, objects and scenes are not combinations of color categories. There is generally no need to categorize or label a color in order to recognize an object or scene. The fact that we *can* categorize colors by labeling them may have implications for how we remember or reason about colors, but because labeling is not an inherent part of a visual process, perhaps we should not expect it to have a significant effect on visual appearance or discrimination.

In sum, there is a growing body of evidence that color terms affect some tasks involving color cognition, but no strong evidence that color terms affect sensory representations. Put another way, the closer a color task is to language, the more likely it is that it will be affected by the terms in one's language. Hue scaling involves explicit judgments about color names, and so it is more likely to be affected by color terms. Memory is more likely to involve linguistic codes than perception, and so memory tasks are more likely to show effects of color terms. For any given task, it might not be immediately obvious whether linguistic codes are used. Future work on the topic would benefit from explicitly modeling the

psychological processes involved in specific color tasks, in order to better understand how color terms affect the various components of cognition such as memory, decisions, and preferences.

Cross-References

- ▶ [Color Categorical Perception](#)
- ▶ [Dynamics of Color Category Formation and Boundaries](#)
- ▶ [Psychological Color Space and Color Terms](#)
- ▶ [World Color Survey](#)

References

1. Pylyshyn, Z.: Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.* **22**(3), 341–365 (1999)
2. Whorf, B.L.: *Language, Thought and Reality*. MIT Press, Cambridge, MA (1956)
3. Harnad, S. (ed.): *Categorical Perception: the Groundwork of Cognition*. Cambridge University Press, Cambridge (1987)
4. Bornstein, M.H., Korda, N.O.: Discrimination and matching within and between hues measured by reaction times: some implications for categorical perception and levels of information processing. *Psychol. Res.* **46**(3), 207–222 (1984)
5. Kuhl, P.K.: Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**(11), 831–843 (2004)
6. The World Color Survey. <http://www1.icsi.berkeley.edu/wcs/>
7. Brown, A.M., Lindsey, D.T., Guckes, K.M.: Color names, color categories, and color-cued visual search: sometimes, color perception is not categorical. *J. Vis.* **11**(12), 1–21 (2011)
8. Davidoff, J., Davies, I., Roberson, D.: Colour categories in a stone-age tribe. *Nature*. **402**(6724), 604–604 (1999)
9. Kay, P., Regier, T., Gilbert, A., Ivry, R.B.: Lateralized Whorf: language influences perceptual decision in the right visual field. In: Minett, J., Wang, W. (eds.) *Language, Evolution, and the Brain*, pp. 261–284. The City University of Hong Kong Press, Hong Kong (2009)
10. Witzel, C., Gegenfurtner, K.R.: Is there a lateralized category effect for color? *J. Vis.* **11**(12), 1–25 (2011)
11. Kay, P., Kempton, W.M.: What is the Sapir-Whorf hypothesis? *Am. Anthropol.* **86**, 65–79 (1984)
12. Webster, M.A., Kay, P.: Color categories and color appearance. *Cognition*. **122**, 375–392 (2012)

-
13. Roberson, D., Davidoff, J.: The categorical perception of colors and facial expressions: the effect of verbal interference. *Mem. Cogn.* **28**, 977–986 (2000)
 14. Witzel, C., Gegenfurtner, K.R.: Categorical facilitation with equally discriminable colors. *J. Vis.* **15**, 1–33 (2015)
 15. Roberson, D., Hanley, J.R., Pak, H.: Thresholds for color discrimination in English and Korean speakers. *Cognition*. **112**, 482–487 (2009)