Universal Foci and Varying Boundaries in Linguistic Color Categories

Terry Regier* **(regier@uchicago.edu)**
Department of Psychology, 5848 S. University Ave.
Chicago, IL 60637 USA

Paul Kay (kay@cogsci.berkeley.edu)
International Computer Science Institute, 1947 Center Street
Berkeley, CA 94704 USA

Richard S. Cook (rscook@icsi.berkeley.edu)
International Computer Science Institute, 1947 Center Street
Berkeley, CA 94704 USA

Abstract

Recent research has questioned the universal basis of color categorization and has instead emphasized cross-linguistic variation in boundaries of color categories. We propose that these cross-linguistically varying boundaries can be predicted from near-universal focal colors within the categories. In support of this proposal, we show that: (1) best example choices for color terms in 110 unwritten languages cluster near the prototypes for English *white, black, red, green, yellow*, and *blue* – we take these 6 points in color space to approximate universal foci; (2) best example choices cluster more tightly across languages than do category centroids; and (3) a computational model can predict color term boundaries from labelings of best examples reasonably well, for several languages, including one that has been taken to counterexemplify universal tendencies in color naming.

Overview

It has long been held that there are universal tendencies in color naming, in that linguistic color categories are organized around universally-shared focal points, or prototypes, in color space. Berlin and Kay (1969; B&K for short) showed that the *best examples* of color terms across a sample of 20 languages seemed to cluster in color space. That study and subsequent work (Kay & McDaniel, 1978; Kay & Maffi, 1999) showed that the most reliable and widespread of these clusters correspond to the six Hering primaries: white, black, red, green, yellow, and blue – suggesting that these points in color space may constitute a universal foundation for color naming. These foci in color space have also appeared to be *cognitively* privileged, in non-linguistic tasks with speakers of languages that have dissimilar color naming systems (Heider, 1972; Heider & Olivier, 1972).

Recently, however, Roberson, Davies, and Davidoff (2000; RDD for short; see also Davidoff, Davies, & Roberson, 1999) turned this universalist account on its head. RDD proposed that color categories are not universal, and are constrained only rather loosely – the most important constraint being that if two points in color space belong to the same linguistic color category, points between them should also belong to that category (p. 395). By implication, the actual location in color space of these categories is not taken to be constrained. RDD suggested moreover that color categories are not organized around universal foci, but are instead determined by naming distinctions made at category *boundaries* – which vary across languages. On their view, foci (best examples) are mere epiphenomena: once categories have been defined by language-determined boundaries, best examples may be derived secondarily as the centers of these already-determined categories (p. 395). The empirical basis for their inversion of the universalist view is: (a) they attempted to replicate, in Berinmo, a Papua New Guinea language, Heider's findings of cognitively privileged status for focal points – and failed to do so; (b) best example choices for some Berinmo color terms are rather diffuse, rather than all falling at or very near the proposed universal foci; (c) Berinmo and English have different boundaries for color terms – and these differing boundaries appear to influence non-linguistic memory for color in speakers of these two languages.

Are the cross-linguistically varying boundaries of color categories determined from universal prototypes (foci) – or are prototypes determined from language-demarcated boundaries? We wished to discriminate between these two proposals, to clarify the broader question of color naming universals. B&K's original universalist findings were based largely on written languages of industrialized societies – thus, the the regularities they found could have resulted from the global spread of industrialization, rather than from genuinely universal forces. Berinmo, in contrast, is a language spoken by an otherwise undocumented group that

* Corresponding author: regier@uchicago.edu

RDD describe as a "stone age culture". In the present study, we relied on a large set of color naming data from unwritten languages of non-industrialized societies, which we compared with data from written languages of industrialized societies, to ensure as best we could that any commonalities we found were genuine universals (Kay & Regier, 2003). We reasoned as follows:

> Prediction 1: If best examples are reflections of the proposed universal foci, then best examples of color terms from unwritten languages should cluster near those locations in color space corresponding to the best examples of English *white, black, red, green, yellow*, and *blue.* This would not be predicted if best examples are instead derived as the centers of cross-linguistically varying category boundaries.

> Prediction 2. If best examples are reflections of universal foci, then best examples should cluster more tightly across languages than do category *centroids* (centers of mass of category extension). This follows since on this view the best examples are universal, while the centroids are affected by category extension, which varies across languages. However, if best examples are derived secondarily as the centers of boundary-defined categories, best examples should not cluster more tightly than centroids (centers).

> Prediction 3: If cross-linguistically varying color category boundaries are projected from privileged foci within those categories, then a computational model should be able to predict color category boundaries from best examples of those categories, for a variety of languages.

We tested these three predictions, in three studies.

Study 1: Universals of color term foci

Do best examples of color terms from unwritten languages cluster near those of English *white, black, red, green, yellow*, and *blue* (and straightforward translations of these terms in other written languages)?

The World Color Survey[1] (WCS; Kay, Berlin, Maffi & Merrifield, 1997; see also Cook, Kay & Regier, in press) collected color naming data from 110 unwritten languages of non-industrialized societies worldwide, from an average of 24 native speakers per language (mode: 25 speakers). Each speaker named, in his or her native language, each of the 330 color chips shown in the stimulus array[2] of Figure 1, and also indicated which chip in the array represented the best example of each color term in the language.

[1] Data at http://www.icsi.berkeley.edu/wcs/data.html

[2] The rows correspond to 10 equally spaced levels of Munsell Value (lightness); the columns correspond to 40 equally spaced Munsell Hues, from R2.5 in column 1 to RP 10 in column 40; the color in each cell corresponds approximately to the maximum available Munsell Chroma for that Hue-Value combination.

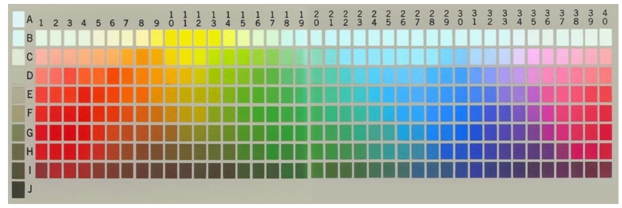

Figure 1: The World Color Survey stimulus array.

We pooled data from all speakers in all languages, and calculated how many WCS best example choices (hits) fell on each chip of the array. The two chips in the array that received the most hits were A0 (2048 hits) and J0 (1988 hits). These two chips lie at the extremes of the leftmost column of achromatic chips. They are the lightest and darkest chips in the array, and align closely with best examples of English *white* and *black*, respectively. In B&K's English color naming data, the best example of *white* was B0, which is one chip away from A0. The B&K stimulus array did not include A0, so it was not available as a possible selection in that study. The best example of English *black* was J0.

The contour plot in Figure 2 shows the number of WCS best example hits that fell on each chip of the stimulus array other than the leftmost column of achromatic chips. The outermost contour represents 100 hits, and each subsequent inner contour represents an increment of 100 hits. The black dots indicate the best examples of the English color terms *red, yellow, green,* and *blue,* provided by one U.S. speaker, as reported by B&K.[3]

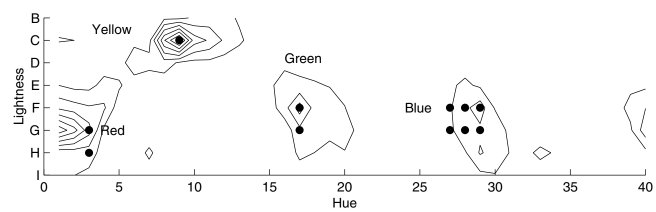

Figure 2. Contour plot of WCS best example choices, compared with best examples of English color terms.

The best examples of named color categories across the 110 unwritten languages of the WCS appear to cluster near or at the best examples of English *white, black, red, green, yellow,* and *blue*. This finding suggests that commonalities in color naming are not restricted to written languages of industrialized societies, and that these 6 regions in color space may reasonably be considered universal foci. This finding would not have been predicted if best examples in these languages are derived from arbitrary language-defined category boundaries.

[3] B&K reported more than one best example choice for several of the English color terms. All best example choices are displayed here.

Concretely, we take the universal foci to be the peaks of the WCS best example distribution: A0 (white: 2048 WCS hits), J0 (black: 1988 hits), C9 (yellow: 752 hits), G1 (red: 668 hits), F17 (green: 351 hits), and F29 (blue: 253 hits).[4]

Figure 3 shows the same distribution of WCS best examples, but this time compared with Berinmo. The numbers in the grid indicate how many speakers located the best example of some Berinmo color term at that position, as reported by RDD. RDD's stimulus array did not include A0, J0 or any other achromatic chips – so the chips in rows B and I, which received many hits, were the closest available approximations to focal white and black respectively. The remaining best example choices peak near the WCS peaks for red, yellow, and green.

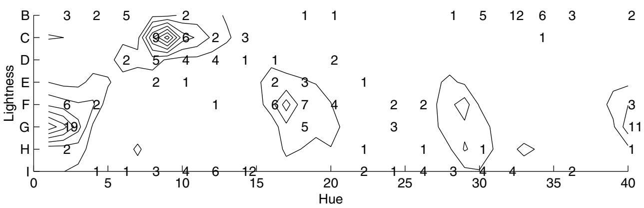

Figure 3. WCS best example choices, compared with those of Berinmo.

If Berinmo foci are similar to those of other languages, and boundaries are determined from foci, we would expect some other languages to also have *boundaries* similar to those of Berinmo. Figure 4 shows the named color categories of Berinmo and Yaminahua (Panoan family, Peru, WCS). Different categories within a language are designated by different colors. Best examples for these Yaminahua color terms peak at A0, J0, C12, F1, and F17 – fairly near the Berinmo peaks. The color category boundaries are also similar across the two languages. The WCS contains several other languages of comparable similarity to Berinmo, such as Colorado (Paezan family, Ecuador), and Iwam (Upper Sepik family, Papua New Guinea).

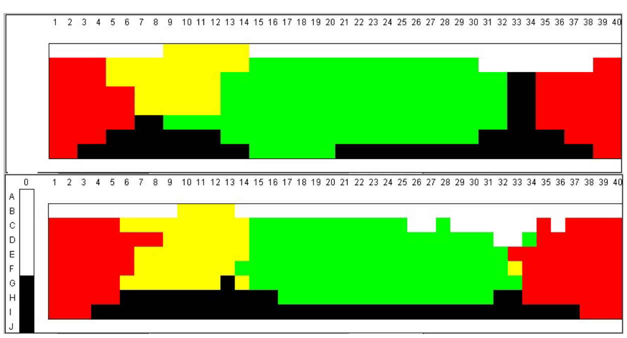

Figure 4: Color categories in Berinmo (upper panel) and Yaminahua (lower panel).[5]

Taken as a whole, this study shows that best examples of unwritten languages, including Berinmo, cluster near the proposed universal foci. Further, as would be predicted if categories are formed around universal foci, Berinmo category *boundaries* are also similar to those of some other languages.

Study 2: Best examples and centroids

Best examples appear to cluster non-randomly (Study 1) – and it has also been shown that the *centroids* of named color categories cluster at rates greater than chance (Kay & Regier, 2003). But do best examples cluster more tightly across languages than centroids do? This is not yet known, and it would be expected if best examples reflect the universal foci around which color categories are organized – because category centroids are sensitive to cross-linguistically varying boundaries, while universal foci are not. In contrast, if best examples are abstracted as the centers of boundary-delineated categories, best examples should cluster no more tightly than category centers – that is, centroids.

Since clustering depends on the idea of distance, we needed a color space with a psychologically meaningful distance metric. We chose CIEL*a*b* space, a 3-dimensional color space which has such a metric, and represented each chip of the WCS array in that space.[6] We calculated the centroid of each color term in each language in the WCS and B&K datasets, following Kay and Regier (2003): for each speaker *s* who used term *t*, we first found the centroid in CIEL*a*b* space of the chips that *s* had named *t*. We then took the average of these speaker centroids for *t*, and coerced it to the chip most similar to it in the stimulus array – this was our representation of the overall term centroid for *t*. We calculated the *focus* of each color term in each language by selecting that chip in the WCS array that received the maximum number of best example choices for that term. If more than one chip received the maximum number, we chose randomly among those chips that had "tied".[7] This left us with two single-point representations for each color term: a centroid and a focus.

We compared the distance separating WCS foci from B&K foci, to the distance separating WCS centroids from B&K centroids. We restricted attention to those terms for which we had both a focus and a centroid – occasionally, one or the other was missing from the data. Then, for each

---

[4] MacLaury (1997: 202) displayed a histogram of WCS focus hits per Munsell hue column, which showed peaks in columns 1, 9, 17, and 29.

[5] The Berinmo data are as reported by RDD, with one qualification: RDD reported naming data only for every other column of the array shown here. We have assumed that each intervening column would have been named exactly as its neighboring column on the right was. The array for Yaminahua was obtained by assigning to each chip the label that was used most often in naming that chip.

[6] See http://www.icsi.berkeley.edu/wcs/data.html under "WCS Mapping Tables" for mappings between CIEL*a*b*, Munsell, and WCS coordinates. The CIEL*a*b* coordinates were obtained using the CMC2 conversion software at www.munsell.com.

[7] This random selection biases our test against finding that foci cluster more tightly than centroids, since focus representations are in part the product of a random process, while centroids are not.

language l in the WCS, we calculated its "centroid separation" (CS_l) from the B&K dataset, as follows: For each term t in l, we found the closest term t^* in each B&K language l^*, and summed the distances – where distances here are defined as CIEL*a*b* distances between centroids, and $c(x)$ stands for "centroid of term x":

$$CS_l = \sum_{t \in l} \sum_{l^* \in BK} \min_{t^* \in l^*} \text{distance}(c(t), c(t^*))$$

We analogously calculated the "focus separation" (FS_l) of each WCS language l from the B&K dataset – this time using distances between foci, rather than between centroids. Here, $f(x)$ stands for "focus of term x":

$$FS_l = \sum_{t \in l} \sum_{l^* \in BK} \min_{t^* \in l^*} \text{distance}(f(t), f(t^*))$$

These calculations gave us, for each language in the WCS, a measure of the distance of its *centroids* from those of B&K, and an analogous measure of the distance of its *foci* from those of B&K. A paired t-test revealed that the focus separation (M=5596.98) was smaller than the centroid separation (M=6391.78), $t(109) = 10.2506$, $p < 0.0001$.

Thus, while centroids do show universal tendencies (Kay & Regier, 2003), best examples appear to show stronger universal tendencies. This pattern is predicted by the hypothesis that best examples reflect the universal structure around which color categories are formed – and it is not predicted by the competing hypothesis that best examples are derived secondarily as the centers of categories that are defined arbitrarily at their boundaries by language.

Study 3: Predicting boundaries from foci

We have so far emphasized universal tendencies in best examples, but the cross-linguistic variation in category boundaries that RDD underscore is very real. Here, we test the hypothesis that cross-linguistically varying boundaries can be predicted from foci by a computational model.

Model

The central concept of the model is that different languages have color terms organized around different *subsets* of the same 6 universal foci (cf. Kay & McDaniel, 1978; Steels & Belpaeme, in press). For instance, some languages have categories organized around only the black, white, and red foci, while other languages make use of other foci as well.[8] In the model, the choice of *which* foci are used affects *where* the category boundaries are predicted to lie.

Although we have seen that the best examples of color terms across languages form clear clusters in color space, there is also evidence that languages vary slightly, but significantly, in the exact location of these clusters (Webster & Kay, in press). Consequently, we examined the best example data for a language to be modeled, and took as "foci" the peaks of the best example distributions for the terms in that language. The model was initialized with these foci, each labeled with the corresponding color term from that language. The model predicted from this the full extension of each of the color terms.

In the model, all color chips of the stimulus array, including the foci, are represented in (3-dimensional) CIEL*a*b* color space. A 3-dimensional Gaussian distribution is centered at each focus to be included in the model. The standard deviations (SDs) in all 3 dimensions of CIEL*a*b* space for a single such "focus distribution" are constrained to be the same – and the single SD associated with each focus is a free parameter of the model. Thus, a model with n foci will have n free parameters governing the widths of the n focus distributions. Given these distributions, and values for the free parameters, the model determines category extensions from foci, by labeling cell i in the stimulus array with the label of the focus distribution that has highest density at that point in color space:

$$label(i) = label(\arg\max_f \ p_f(i))$$

Here, f ranges over foci in the model, and $p_f(i)$ represents the density function of the focus distribution for f at the point in CIEL*a*b* space corresponding to chip i of the stimulus array. Once all points have been labeled, the results may be mapped back to, and displayed in, the WCS stimulus array.

The model's free parameters (SDs) were fit to data by a coarse-to-fine search. Each free parameter initially ranged from 1-15, with step size 2 (less systematic searches of broader ranges did not yield better results). We noted those parameter values that yielded the maximum number of chips correctly classified. A series of successively finer searches was then initiated, with the ith search ranging over values one i-1 step size above and below the optimum values for search i-1. The decreasing step sizes were 2, 1, .5, .2, .1. The best fit for the finest of these searches was taken as the overall fit of the model to the data.

Berinmo

Given the Berinmo best example data shown in Figure 3, we analyzed Berinmo as having 3 chromatic foci: red, yellow, and green (located at the 3 chromatic peaks in the Berinmo best example data: G2, C8, and F18 – each is one chip away from a WCS peak). We also assumed that the remaining two Berinmo terms were focused at universal white (A0) and black (J0). We fit the model, with the above 5 foci, to RDD's Berinmo naming data. These data are shown (again) in the upper panel of Figure 5, while the lower panel shows the model's fit to the data. The best-fitting parameter values were: SD(black) = 5.1, SD(white) = 3.2, SD(red) = 7.1, SD(yellow) = 8.6, SD(green) = 10.1. 90% of chips were correctly predicted. Several qualitative features of the Berinmo data are also captured by the model: (1) a category focused near yellow that extends into what we would call *green* in English, (2) a category that "encompasses much of

[8] Some languages, like English, appear to also use foci other than the six discussed here. These languages are beyond the scope of this paper.

green, blue, and blue-purple" (RDD, p. 377), focused one chip away from universal green, (3) a white/light category that extends farthest downward around columns 28-38, and (4) a black/dark category that extends somewhat up into brown and purple – although in the simulation it does not extend as far as in the data.

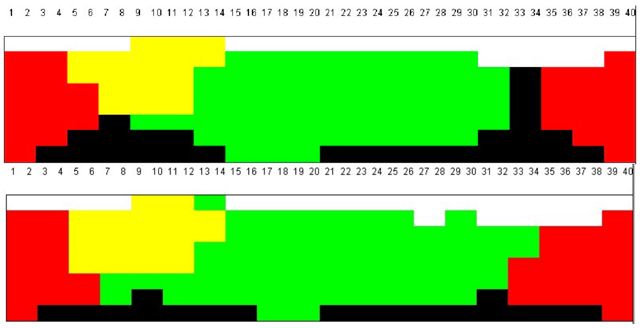

Figure 5: Berinmo naming data (top panel), and model fit to those data, based on Berinmo foci (bottom panel). 90% of chips are correctly labeled.

We also fit the model to Berinmo in a slightly different manner: by assuming that the 5 foci were located at the black, white, red, yellow, and green *WCS peaks* from Study 1, rather than the Berinmo foci which fall 1 chip away from them. The best-fitting parameter values were: SD(black) = 5.5, SD(white) = 3.5, SD(red) = 6.5, SD(yellow) = 8.5, SD(green) = 12.7. This time, with foci not drawn from the language itself, the fit was not quite as good: 84% of chips were correctly classified. These results suggest that the rather small differences in the locations of foci across languages may account for some cross-linguistic variation in category boundaries.

Other languages

A natural objection to the above demonstration is that the model was fit directly to the Berinmo data: with 5 free parameters, a reasonable fit may be unsurprising.[9] A stronger test of the model would be to see whether the parameter values obtained by fitting one language can also predict boundaries in a rather different set of languages.

We considered the subset of languages in the WCS that have color terms centered in three foci: black (near J0), white (near A0), and red (near G1). There are four such languages in the WCS, all spoken in west Africa, and with naming data shown in the top four panels of Figure 6 – from the top down: Bete (Kru, Ivory Coast), Ejagam (Bantoid, Nigeria and Cameroon), Wobé (Niger-Congo, Ivory Coast), and Yacouba (Niger-Congo, Ivory Coast). The category boundaries of these languages are similar to each other – and dissimilar from Berinmo.

We retained the black, white, and red distributions in the model, centered at the WCS peaks,[10] with the SDs fixed at the values obtained by fitting the *Berinmo* data – and removed the yellow and green focal distributions. We then predicted category boundaries from foci using this 3-focus model. The results are shown in the bottom panel of Figure 6. Qualitatively, the boundaries are similar to those found in the 3-focus languages of the WCS. Table 1 displays quantitative measures of fit.

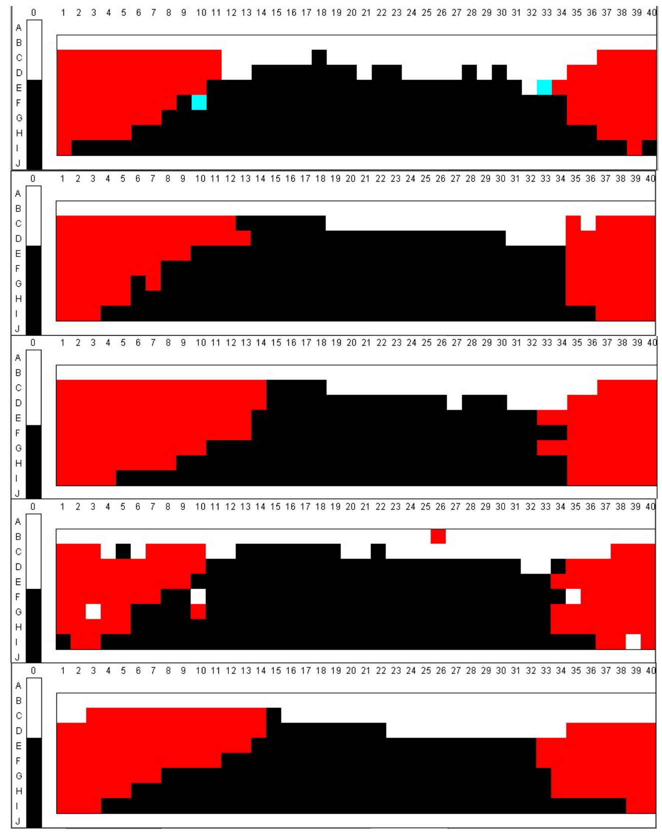

Figure 6: Top four panels show naming data from four 3-focus WCS languages (see text). Bottom panel shows model prediction given 3 WCS foci, with parameters fit to Berinmo naming data.

Table 1: Model fit to 3-focus languages, using WCS foci. Free parameters were fit to Berinmo naming data.

Language	**Percent correct**
Bete	89%
Ejagam	88%
Wobé	90%
Yacouba	84%

Thus, the structures and parameter values that provide a reasonably good fit to the Berinmo data also provide a fairly

[9] Still, in other simulations, we have shown that there are many possible (but unattested) color categorization patterns that this model cannot fit well.

[10] We chose WCS peaks, rather than peaks from Berinmo or the 3-focus languages, since we wished to approximate universal foci in this simulation.

good fit to languages from a different part of the world, with color category boundaries unlike those of Berinmo. Any claims concerning the overall adequacy of this model will have to await the results of far more comprehensive tests. However, for now, the model does appear to support the idea that cross-linguistically varying category boundaries can be predicted fairly well from near-universal foci.

Discussion

We take these results to cast into question RDD's proposal that color categories are demarcated at their boundaries by language in an only loosely constrained fashion, and that best examples are epiphenomena of this process. Instead, we view these results as supporting a universal tendency for the named color categories of languages to be based on favored percepts selected from restricted regions of color space. The degree to which these universally favored regions are based on color appearance (Kay & Maffi, 1999), on universal statistical tendencies in the distribution of reflective surfaces in the environment (Yendrikhovskij, 2001), on universal properties of ambient light sources (Shepard, 1992), on the topography of perceptual color space (Jameson & D'Andrade, 1997), or on socio-linguistic negotiation among speakers (Steels & Belpaeme, in press), cannot be assessed with any degree of certainty at this time. Possibly all these factors, and perhaps others, play a role. It is similarly not yet clear to what extent cross-language variation in the precise location of foci may affect the boundaries of color categories – or even to what extent it is legitimate to approximate foci as points at all, rather than (possibly somewhat irregularly shaped) areas. That named color categories in the world's languages are based to a considerable degree on such favored regions of color space can, however, be asserted with some confidence.

At the same time, there is by now considerable evidence that color category boundaries can affect non-linguistic color cognition (e.g. Kay & Kempton, 1984; RDD; Witthoft, Winawer, Wu, Frank, Wade, & Boroditsky, 2003). We take our present findings to be compatible with such Whorfian results, provided one allows that the placement of category boundaries is itself constrained by universal forces.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants Nos. 0130420 and 0418283.

References

Berlin, Brent and Paul Kay (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley, California: University of California Press.

Cook, Richard S., Paul Kay, and Terry Regier (in press). The World Color Survey database: History and use. In Henri Cohen and Claire Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science*. Elsevier.

Davidoff, Jules, Ian Davies, and Debi Roberson (1999). Colour categories in a stone-age tribe. *Nature, 398*, 203-204.

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology 93*, 10-20.

Heider, E. R. and D. C. Olivier (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology, 3*, 337-354.

Jameson, Kimberly and Roy G. D'Andrade (1997). It's not really Red, Green, Yellow, Blue: An inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language*. Cambridge, England: Cambridge University Press.

Kay, Paul, Brent Berlin, Luisa Maffi, and William Merrifield (1997). Color naming across languages. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language*, pp. 21-56. Cambridge University Press, Cambridge, England.

Kay, Paul and Willett Kempton (1984). What is the Sapir-Whorf Hypothesis? *American Anthropologist, 86*, 65-79.

Kay, Paul and Luisa Maffi (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist, 101*, 743-760.

Kay, Paul and Chad McDaniel (1978). The linguistic significance of the meanings of basic color terms. *Language, 54*, 610-646.

Kay, Paul and Terry Regier (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences, 100*, 9085-9089.

MacLaury, Robert E. (1997). Ethnographic evidence of unique hues and elemental colors. *Behavioral and Brain Sciences, 20*, 202-203.

Roberson, Debi, Ian Davies and Jules Davidoff (2000). Colour categories are not universal: Replications and new evidence from a stone age culture. *Journal of Experimental Psychology: General, 129*, 369-398.

Shepard, Roger N. (1992) The perceptual organization of colors. In J. Barkow, L. Cosmides, and J. Tooby (Eds.), *The Adapted Mind*. Oxford : Oxford University Press.

Steels, Luc and Tony Belpaeme (in press) Coordinating perceptual categories through language: A case study for colour. *Behavioral and Brain Sciences.*

Webster, Michael A. and Paul Kay (in press). Individual and population differences in focal colors. In Robert E. MacLaury, Galina V. Paramei and Don Dedrick (Eds.), *The Anthropology of Color*. Amsterdam: John Benjamins.

Witthoft, Nathan, Jonathan Winawer, Lisa Wu, Michael Frank, Alex Wade, and Lera Boroditsky (2003). Effects of language on color discriminability. In Richard Alterman and David Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* [CD].

Yendrikhovskij, Sergej N. (2001). Computing color categories from statistics of natural images. *Journal of Imaging Science and Technology, 45*, 409-417.