Variations in color naming within and across populations

Michael A. Webster[a] and Paul Kay[b]

[a]Department of Psychology, University of Nevada – Reno, Reno, NV 89557;

[b]*International Computer Science Institute and University of California – Berkeley,*

Berkeley, CA 94704.

mwebster@unr.nevada.edu

kay@icsi.berkeley.edu

http://www.icsi.berkeley.edu/~kay/

Abstract: The simulations of Steels & Belpaeme suggest that communication could lead

to color categories that are closely shared within a language and potentially diverge

across languages. We argue that this is opposite of the patterns that are actually observed

in empirical studies of color naming. Focal color choices more often exhibit strong

concordance across languages while also showing pronounced variability within any

language.

Steels & Belpaeme use theoretical simulations to explore the potential role of

physiological, environmental, and cultural (linguistic) constraints on the acquisition of

shared color categories. Although their stated aim is to identify principles that could

guide the design of communication among artificial intelligence systems, they emphasize

that the results are also relevant for understanding color categorization in human observers. Our commentary focuses on the extent to which the trends they observe are evident in actual studies of color naming.

In Steels & Belpaeme's simulations, whether or not a factor provides a loose or tight constraint is evaluated by measuring the variance in color categories across observers. In all cases, they find the variance to be greater for agents drawn from separate populations than for those drawn from the same population, yet this difference becomes dramatic when the categories are learned through language, in which case, the within-group variance approaches zero. This, in theory, points to a strong potential for cultural relativity in color naming.

What are the patterns of variance in empirical measures of color naming? There are two striking patterns. First, there are strong universal tendencies across languages. These tendencies were originally suggested by Berlin and Kay (1969) and have been confirmed by Kay and Regier (2003) in a recent analysis of the World Color Survey (WCS), which provides color-naming responses from an average of 24 primarily monolingual speakers from each of 110 unwritten languages. Specifically, they showed that the centroids of color-naming responses for different languages exhibit much stronger clustering than would be predicted by chance. This is qualitatively consistent with Steels & Belpaeme's analyses, showing that physiological and/or environmental constraints can support some degree of consistency among speakers. Whether it is quantitatively consistent could potentially be evaluated by applying the authors' variance metric to the WCS data (which

is available on-line at http://www.icsi.berkeley.edu/wcs/data.html). This might allow one to assess whether different languages show more concordance in color categories than would be expected from their models of physiological and environmental factors. Without such comparisons, it is difficult to interpret the relevance for human behavior of the values they derive from simulations.

The second prominent property of actual color-naming data is the pronounced variation among speakers of the same language. Individual differences in unique hue and focal color choices have been widely documented, though their causes remain poorly understood (Webster et al. 2000). For example, the wavelengths that individuals select for unique green within a linguistically homogeneous group span a range of more than 80 nm; these variations are in fact so large that the same wavelength might be chosen as unique green by one observer and unique yellow or blue by another (Kuehni 2004). Individual differences in focal color choices remain large for more naturalistic spectra like the Munsell chips and represent another obvious feature of the WCS data (as well as for most other data sets on color naming). Moreover, comparable differences persist even when the samples are restricted to individuals who select colors with the highest reliability (Webster et al. 2000). In sum, in actual measures of color naming, as contrasted to simulations, within-group variance is very large.

This fact appears difficult to reconcile with the minimal variance predicted by Steels & Belpaeme to arise from adding communication to the simulated agents. Actual agents do not show the close agreement that language could potentially support. As an illustration

of this, Table 1compares the average within-language variance to the variance in mean foci across languages for "red," "green," "blue," or "yellow" terms for the WCS respondents, based on an analysis by Webster and Kay (in press). (These are calculated from the raw distances in the Munsell palette for the Hue and Value dimensions separately.) For each language, terms corresponding to the English terms were determined by finding the focal choices for consensus terms closest to the English foci. For "red" and "yellow" these correspondences are obvious. For example, the means for the "red" and "yellow" clusters are separated by approximately ten times the cluster standard deviations, with only one language exhibiting a consensus term nearer to the intermediate focal point for English "orange." This finding echoes the consistent clustering demonstrated by Kay and Regier (2003). For green and blue the clustering is less obvious because many of the WCS languages apply a common "grue" term to this region. The values shown are thus restricted to the subset of languages that have both terms. Mean foci across languages vary much less than individual foci within languages. This suggests that a common language imposes only a weak constraint, and a difference in language produces relatively little divergence.

Another finding that may argue against a strong constraint of language on human color categories is that individual differences in focal choices for binary hues (e.g., blue-green or yellow-green) are not obviously distinct from the differences measured for primary hues (e.g., blue or yellow or green). Malkoc et al. (2002) in fact found less variation in "focal" blue-green than in the unique hues blue and green. That is, English speakers were more consistent at selecting the boundary between blue and green than at choosing either

primary category's best example, even though there is no basic term targeting this boundary, and, as noted, many other languages do not have separate words for these categories.

Admittedly, the loci of color categories do vary significantly across different language groups, and there are both extreme (Davidoff et al. 1999) as well as more subtle examples of these differences (Webster et al. 2002). For example, Webster and Kay analyzed whether differences between the average foci for the WCS languages were larger than predicted by random sampling across languages; and as Table 1 shows, differences were significant for all terms. Nevertheless, as indicated earlier, the within language variances are much higher than the variances in mean foci between languages. As Steels & Belpaeme show, interlanguage differences can arise from many sources. The question remains as to exactly what degree the existence of different languages is an actual contributing factor to the total interpersonal variation in color naming.

References

Berlin, B. & Kay, P. (1969) *Basic color terms: Their universality and evolution*. University of California Press.

Davidoff, J., Davies, I. & Roberson, D. (1999) Colour categories in a stone-age tribe. *Nature* 398:203–04.

Kay, P. & Regier, T. (2003) Resolving the question of color naming universals. *Proceedings of the National Academy of Science* 100:9085–89.

Kuehni, R. G. (2004) Variability in unique hue selection: A surprising phenomenon. *Color Research and Application* 29:158–62.

Malkoc, G., Kay, P. & Webster, M. A. (2002) Individual differences in unique and binary hues. *Journal of Vision* 2:32a.

Webster, M. A. & Kay, P. (in press) Individual and population differences in focal colors. In: *The anthropology of color*, ed. R. L. MacLaury, G. Paramei & D. Dedrick. Benjamins.

Webster, M. A., Miyahara, E., Malkoc, G. & Raker, V. E. (2000) Variations in normal color vision. II. Unique hues. *Journal of the Optical Society of America A* 17:1545–55.

Webster, M. A., Webster, S. M., Bharadwadj, S., Verma, R., Jaikumar, J., Madan, G. & Vaithilingam, E. (2002) Variations in normal color vision. III. Unique hues in Indian and United States observers. *Journal of the Optical Society of America A* 19:1951–62.

Table 1. Average variance in individual foci within a WCS language compared to the variance of mean foci between languages, computed for the hue or lightness of "red," "green," "blue," or "yellow." *F*-tests compare the between-language variance to the variance predicted by randomly sampling speakers of different languages. The hue scale runs from 1 = Munsell 2.5R, in 40 steps, to 40 = Munsell 10RP. The lightness scale is Munsell Value.

term	#		focal hue						focal lightness				
			mean	var	pred	F	p		mean	var	pred	F	p
r	103		1.77	.46	.25	1.81	<.002		4.25	.095	.040	2.41	< e-5
g	73		18.9	3.01	.96	3.16	< e-8		4.74	.41	.099	4.12	< e-10
b	50		27.7	2.45	.93	2.56	< e-5		4.30	.46	.093	4.84	< e-10
y	86		9.46	.65	.31	2.13	<.0002		7.79	.13	.038	3.38	< e-8