Focal colors are universal after all

Terry Regier[†‡§], Paul Kay[‡¶], and Richard S. Cook[¶]

[†]Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637; and [¶]International Computer Science Institute, 1947 Center Street, Berkeley, CA 94704

Contributed by Paul Kay, April 21, 2005

It is widely held that named color categories in the world's languages are organized around universal focal colors and that these focal colors tend to be chosen as the best examples of color terms across languages. However, this notion has been supported primarily by data from languages of industrialized societies. In contrast, recent research on a language from a nonindustrialized society has called this idea into question. We examine color-naming data from languages of 110 nonindustrialized societies and show that (*i*) best-example choices for color terms in these languages cluster near the prototypes for English *white*, *black*, *red*, *green*, *yellow*, and *blue*, and (*ii*) best-example choices cluster more tightly across languages than do the centers of category extensions, suggesting that universal best examples (foci) may be the source of universal tendencies in color naming.

Whorf | linguistic relativity | Berinmo | best example | categories

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 (1) 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 (2, 3) showed that the most reliable and widespread of these clusters correspond to the six Hering primaries (white, black, red, yellow, green, and blue), suggesting that these points in color space may constitute a universal foundation for color naming. These foci in color space have also seemed to be cognitively privileged in nonlinguistic tasks with speakers of languages that have dissimilar color-naming systems (4, 5).

Recently, however, Roberson and coworkers (ref. 6; see also ref. 7) turned this universalist account on its head. They proposed that "color categories [are] a function of cultural experience and only, at most, loosely constrained by the default neural organization." They discuss explicitly only one constraint: "The most important constraint would be that similar items (as defined by perceptual discrimination) are universally grouped together. Thus, no language would exhibit categories that include two areas of color space but excludes [sic] an area in between them" (ref. 6, p. 395). By implication, the actual location in color space of these categories is taken to be unconstrained. Roberson and coworkers suggest, 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. "We argue that color categories are developed from demarcation at boundaries…although the central tendency of exemplars can be extracted at a later stage" (ref. 6, p. 395). "The development of focal stimuli…may represent a second phase of categorization…[o]nce a category has been delineated at the boundaries, exposure to exemplars may lead to the abstraction of a central tendency so that observers behave as if their categories have prototypes" (ref. 6, p. 395). The empirical bases for their inversion of the universalist view are: (*i*) they attempted to replicate, in Berinmo (a Papua New Guinea language), Heider's (4) findings of cognitively privileged status for focal points and failed to do so; (*ii*) best-example choices for some Berinmo color terms are rather diffuse rather than all falling at or very near the proposed universal foci; and (*iii*) where Berinmo and English have different boundaries for color terms, the differing boundaries seem to influence nonlinguistic memory for color in speakers of these two languages. These are unquestionably interesting empirical findings, but it does not follow from them that, as Roberson and coworkers suggest, the boundaries of color categories are determined in each language independently by local linguistic convention and that focal colors merely represent the centers of these culturally determined and culturally varying categories. In this article we show that their conclusion fails empirically as well.

Berinmo is a language spoken by an otherwise undocumented group that Roberson *et al*. (6) describe as a "stone-age culture." In contrast, the original universalist findings of Berlin and Kay (1) were based largely on written languages of industrialized societies; thus, the regularities that Berlin and Kay found could have resulted from the global spread of industrialization rather than from genuinely universal forces. In the present study, as in an earlier one (8), we relied on a large set of color-naming data from unwritten languages of nonindustrialized societies that we compared with data from written languages of industrialized societies to ensure as best as we could that any commonalities we found were genuine universals. The earlier study (8) provided the first objective demonstration of universal tendencies in color naming[‖] but did not address the possible role of universal foci in those tendencies. We sought to do so in the present study and reasoned with the following predictions:

1. If best examples are reflections of the proposed universal foci, then best examples of color terms from languages of nonindustrialized societies 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 derived instead as the centers of categories that are defined at their boundaries by local linguistic convention.
2. If best examples are reflections of universal foci, then best examples should cluster more tightly across languages than do the centers of category extensions, because category extension is known to vary across languages. However, if best examples are derived secondarily as the centers of categories that are defined at their boundaries by language, best examples should not cluster more tightly than the centers of category extensions, because on this view, best examples are category centers.

We tested these predictions in two studies.

Abbreviation: WCS, World Color Survey.

[‡]T.R. and P.K. contributed equally to this work.

[§]To whom correspondence should be addressed. E-mail: regier@uchicago.edu.

[‖]There were some minor inaccuracies in the data processing on which these earlier results were based. We have since then rerun the analyses and obtained the same qualitative results.

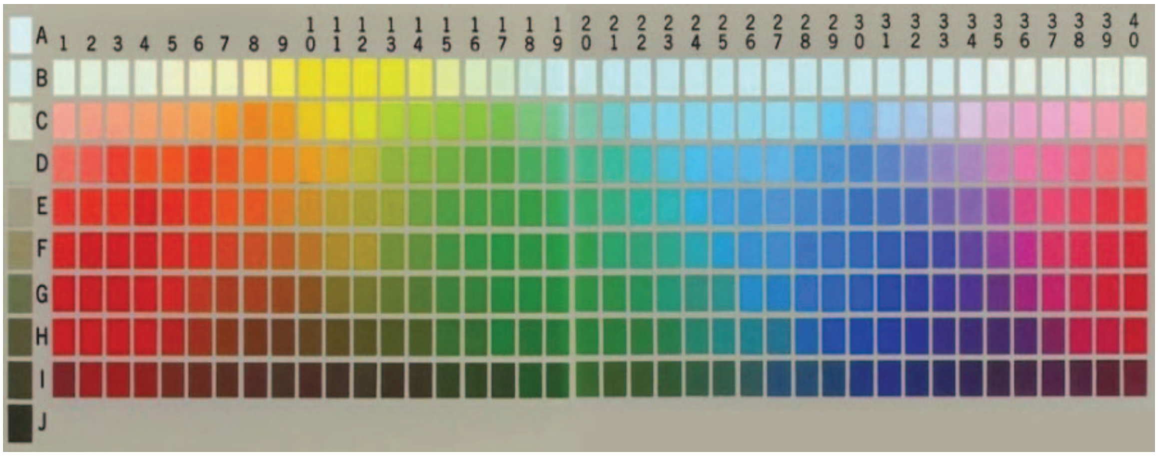

Fig. 1. The WCS stimulus array. The rows correspond to 10 levels of Munsell value (lightness), and the columns correspond to 40 equally spaced Munsell hues, from R2.5 in column 1 to RP10 in column 40. The color in each cell corresponds approximately to the maximum available Munsell chroma for that hue–value combination.

Study 1: Universals of Color-Term Foci

Do best examples of color terms from languages of nonindustrialized societies 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 (WCS; data are available at www.icsi.berkeley.edu/wcs/data.html) (ref. 9; see also ref. 10) collected color-naming data from 110 unwritten languages of nonindustrialized societies worldwide from an average of 24 native speakers per language (mode: 25 speakers). The names of the WCS languages with the families to which they belong and the country in which each was encountered are listed in Table 1. Each speaker named, in his or her native language, each of the 330 color chips shown in the stimulus array of Fig. 1 (we refer to these data as "naming data"). Each speaker also indicated which chip (or sometimes chips) in the array represented the best example, or focus, of each color term in the language (which we refer to as "focus data"). Kay and Regier (8) showed universal tendencies in the naming data only; here, we ask whether these universal naming tendencies may stem from universals in the foci.

We pooled all of the focus data from all speakers in all languages of the WCS and calculated how many best-example choices (hits) fell on each chip of the array. The two chips in the array that received the most hits were A0 (2,048 hits) and J0 (1,988 hits). These two chips lie at the extremes of the left-most 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 the Berlin and Kay English color-naming data, the best example of *white* was B0, which is one chip away from A0. Their 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 Fig. 2 shows the number of WCS best-example hits that fell on each chip of the stimulus array other than the left-most column of achromatic chips. The outermost contour of each cluster 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 Berlin and Kay).

The best examples of named color categories across the 110 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 the best examples of color categories are not restricted to languages of industrialized societies and that these six regions in color space may reasonably be considered to be universal foci. This finding would not have been predicted if best examples in these languages were derived from language-defined category boundaries. Concretely, we take the universal foci to be the peaks of the WCS best-example distribution: A0 (white: 2,048 WCS hits), J0 (black: 1,988 hits), G1 (red: 668 hits), C9 (yellow: 752 hits), F17 (green: 351 hits), and F29 (blue: 253 hits).[††]

[††]MacLaury (ref. 11, p. 202) displayed a histogram of WCS focus hits per Munsell hue column, which showed peaks in columns 1, 9, 17, and 29.

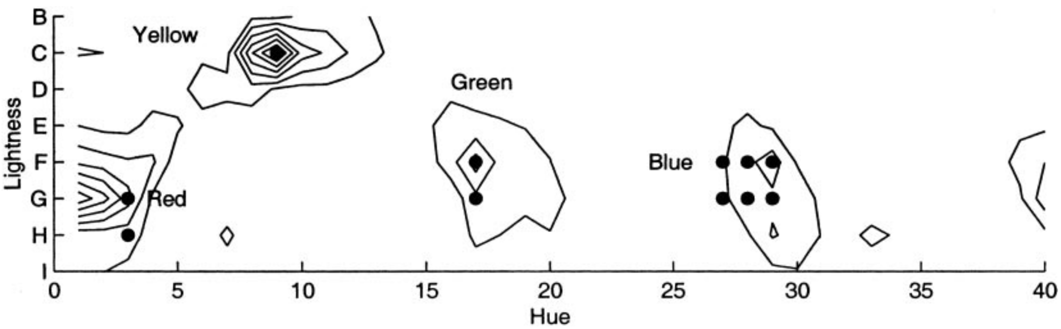

Fig. 2. Contour plot of WCS best-example choices compared with best examples of English color terms. Berlin and Kay reported more than one best-example choice for several of the English color terms; all best-example choices are displayed here.

Table 1. WCS languages, families, and countries where encountered

Index	Language	Family	Country where encountered
1	Abidji	Kwa	Ivory Coast
2	Agarabi	Trans-New Guinea	Papua New Guinea
3	Agta	Austronesian	Philippines
4	Aguacatec	Mayan	Guatemala
5	Amarakaeri	Arawakan	Peru
6	Ampeeli	Angan	Papua New Guinea
7	Amuzgo	Oto-Manguean	Mexico
8	Angaatiha	Angan	Papua New Guinea
9	Apinayé	Macro-Ge	Brazil
10	Arabela	Zaparoan	Peru
11	Bahinemo	Sepik Hill	Papua New Guinea
12	Bauzi	Geelvink Bay	Indonesia
13	Berik	Trans-New Guinea	Indonesia (Irian Jaya)
14	Bété	Kru	Ivory Coast
15	Bhili	Indic	India
16	Buglere	Chibchan	Panama
17	Cakchiquel	Mayan	Guatemala
18	Campa	Arawakan	Peru
19	Camsa	Camsa	Columbia
20	Candoshi	Jivaroan	Peru
21	Cavineña	Tacanan	Bolivia
22	Cayapa	Barbacoan-Paezan	Ecuador
23	Chácobo	Panoan	Bolivia
24	Chavacano (Zamboangueño)	[creole]	Philippines
25	Chayahuita	Chayahuita (Jivaroan?)	Peru
26	Chinantec	Oto-Manguean	Mexico
27	Chiquitano	Macro-Ge	Bolivia
28	Chumburu	Chumburung	Ghana
29	Cofán	Chibchan	Ecuador
30	Colorado	Barbacoan	Ecuador
31	Cree	Algonquian	Canada
32	Culina	Arauan	Peru, Brazil
33	Didinga	Nilo-Sarahan	Sudan
34	Djuka	[creole]	Surinam
35	Dyimini	Gur	Ivory Coast
36	Ejagam	Bantoid	Nigeria, Cameroon
37	Ese Ejja	Tacanan	Bolivia
38	Garífuna (Black Carib)	[creole]	Guatemala
39	Guahibo	Arawakan	Colombia
40	Guambiano	Paezan	Columbia
41	Guarijío	Uto-Aztecan	Mexico
42	Guaymí (Ngäbere)	Chibchan	Panama
43	Gunu	Bantoid	Cameroon
44	Halbi	Indic	India
45	Huastec	Mayan	Mexico
46	Huave	Huavean	Mexico
47	Iduna	Oceanic	Papua New Guinea
48	Ifugao	Austronesian	Philippines
49	Iwam	Upper Sepik	Papua New Guinea
50	Jicaque	Jicaque	Honduras
51	Kalam	East New Guinea Highlands	Papua New Guinea
52	Kamano-Kafe	East New Guinea Highlands	Papua New Guinea
53	Karajá	Macro-Ge	Brazil
54	Kemtuik	Nimboran	Indonesia (Irian Jaya)
55	Kokni (Kokoni)	Indic	India

Roberson *et al*. (6) report five color terms for Berinmo, focused at or near black, white, red, yellow, and green. The term focused at green includes blue and purple. They emphasize the fact that the term focused at yellow includes yellowish greens and in this way differs from English. Berinmo, however, is not unlike several WCS five-color-term languages in these respects (12). Fig. 3 shows the same distribution of WCS best examples as Fig. 2, this time compared with Berinmo. The size of the dots in the grid indicates how many Berinmo speakers located the best example of some color term at that position, as reported by Roberson *et al*.: a small dot denotes one to three best-example choices ("hits"), a medium-sized dot denotes four to five hits, and a large dot denotes six or more hits. The Roberson *et al*. stimulus array did not include A0, J0, or any

Table 1. (continued)

Index	Language	Family	Country where encountered
56	Konkomba	Gur	Ghana
57	Kriol	[creole]	Australia
58	Kuku-Yalanji	Pama-Nyungan	Australia
59	Kuna	Cuna (?), Chibchan (?)	Panama
60	Kwerba	Dani-Kwerba	Indonesia (Irian Jaya)
61	Lele	East Chadic	Chad
62	Mampruli	Gur	Ghana
63	Maring	Trans-New Guinea	Papua New Guinea
64	Martu-Wangka	Martu-Wangka	Australia
65	Mawchi	Indic	India
66	Mayoruna	Panoan	Peru
67	Mazahua	Oto-Manguean	Mexico
68	Mazatec	Oto-Manguean	Mexico
69	Menye	Angan	Papua New Guinea
70	Micmac	Algonquian	Canada
71	Mikasuki	Muskogean	United States
72	Mixtec	Oto-Manguean	Mexico
73	Mundu	Adamawa-Ubangi	Sudan
74	Múra-Pirahá	Pirahá	Brazil
75	Murle	Nilo-Sarahan	Sudan
76	Murrinh-Patha	Murrinh-Patha	Australia
77	Nafaanra	Gur	Ghana
78	Nahuatl	Uto-Aztecan	Mexico
79	Ocaina	Witotoan	Peru
80	Papago (O' odham)	Uto-Aztecan	United States, Mexico
81	Patep	Austronesian	Papua New Guinea
82	Paya	Chibchan	Honduras
83	Podopa	Trans-New Guinea	Papua New Guinea
84	Saramaccan	[creole]	Surinam
85	Seri	Hokan	Mexico
86	Shipibo	Panoan	Peru
87	Sirionó	Tupi	Bolivia
88	Slave	Athabaskan	Canada
89	Sursurunga	Austronesian	Papua New Guinea
90	Tabla	Trans-New Guinea	Indonesia (Irian Jaya)
91	Tacana	Tacanan	Bolivia
92	Tarahumara (Central dialect)	Uto-Aztecan	Mexico
93	Tarahumara (Western dialect)	Uto-Aztecan	Mexico
94	Tboli	Austronesian	Philippines
95	Teribe	Chibchan	Panama
96	Ticuna	Ticuna	Peru
97	Tifal	Trans-New Guinea	Papua New Guinea
98	Tlapanec	Subtiaba-Tlapanec	Mexico
99	Tucano	Tucanoan	Colombia
100	Vagla	Gur	Ghana
101	Vasavi	Indic	India
102	Waorani (Auca, Huao)	Waorani	Ecuador
103	Walpiri	Pama-Nyungan	Australia
104	Wobé	Niger-Congo	Ivory Coast
105	Yacouba	Dan	Ivory Coast
106	Yakan	Austronesian	Philippines
107	Yaminahua	Panoan	Peru
108	Yucuna	Arawakan	Colombia
109	Yupik	Eskimo-Aleut	United States
110	Zapotec	Oto-Manguean	Mexico

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. Roberson *et al*. collected data for only even-numbered columns in the array; thus, the red, yellow, and green universal foci (G1, C9, and F17, respectively) were not themselves available as selections to Berinmo participants.

Study 2: Best Examples and Category Extension

We asked next whether best examples cluster more tightly across languages than do the centers of category extensions. This pattern would be expected if best examples reflect universal foci against a background of cross-linguistically varying category extensions. However, it would not be predicted if best examples are abstracted instead as the centers of categories defined at

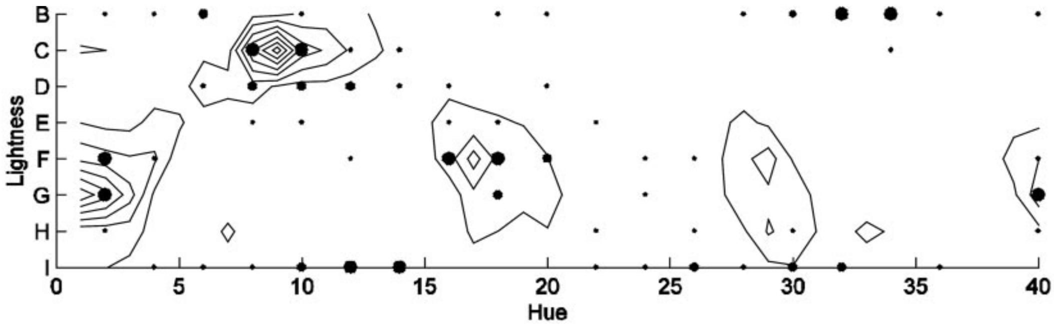

Fig. 3. WCS best-example choices, compared with those of Berinmo.

their boundaries by linguistic convention, because on this latter view, best examples are category centers and will cluster only as tightly as those centers.

To ensure that any regularities we found reflected universal tendencies rather than characteristics of a particular data set, we compared across the WCS and Berlin and Kay data sets, thus comparing languages of societies that were dissimilar with regard to both literacy and level of technology. Specifically, we compared the degree to which centers of category extension in the WCS lie near those in the Berlin and Kay data set with the degree to which WCS best examples (foci) lie near those of Berlin and Kay.

We needed a color space with a psychologically meaningful distance metric. We chose CIEL*a*b* space, a 3D color space that has such a metric, and represented each chip of the WCS array in that space.[‡‡] We calculated the center of category extension of each color term in each language in the WCS and Berlin and Kay data sets by following Kay and Regier (8): 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* (based on naming data, not focus data). 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 naming centroid (henceforth referred to as just "centroid") for term *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 (based on focus data, not naming data). If more than one chip received the maximum number, we chose randomly among those chips that had tied.[§§] This choice left us with two single-point representations for each color term: a centroid and a focus. We restricted attention to those terms for which we had both a centroid and a focus; occasionally, one or the other was missing from the data. Then, for each language l in the WCS, we calculated its centroid separation (CS_l) from the Berlin and Kay data set as follows: For each term t in l, we found the closest term t^* in each Berlin and Kay 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 Berlin and Kay data set, 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 naming centroids from those of Berlin and Kay and an analogous measure of the distance of its foci from those of Berlin and Kay. A paired t test revealed that the focus separation (M = 5,596.98) was smaller than the centroid separation (M = 6,391.78) [$t(109) = 10.2506$; $P < 0.0001$].

Thus, best examples of color categories cluster more tightly across languages of industrialized and nonindustrialized societies than do the centers of those categories' extensions. 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 at their boundaries by the local culture.

Discussion

We take these results to refute the proposal by Roberson *et al*. (6) that color categories are demarcated at their boundaries by local linguistic convention 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 in languages of both industrialized and nonindustrialized societies. The degree to which these universally favored regions are based on color appearance (3), universal statistical tendencies in the distribution of reflective surfaces in the environment (13), universal properties of ambient light sources (14), the topography of perceptual color space (15), or sociolinguistic negotiation among speakers (16) cannot be assessed with any degree of certainty at this time. It is possible that all these factors, and perhaps others, play a role. Similarly, it is not yet clear to what extent cross-language variation in the precise location of foci (17) 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, however, can be asserted with some confidence. At the same time, Roberson *et al.* (6), among others (e.g., refs. 18 and 19), have presented considerable evidence that the cross-linguistically varying boundaries of linguistic color categories can affect nonlinguistic color cognition. We take our present findings to be compatible with such Whorfian results, provided one allows that the variation of category boundaries itself is constrained by universal forces.

[‡‡]See 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 by using CMC2 conversion software (available at www.munsell.com). The L*a*b* space was constructed by the Comité Internationale d'Eclairage so that the just-noticeable difference calculated in any direction from any point is uniform throughout.

[§§]This random selection biases our test against finding that foci cluster more tightly than centroids, because focus representations are in part the product of a random process, whereas centroids are not. The reason we used modal focal choices rather than centroids of focal choices is to avoid any misleading averaging of terms that receive best-example responses in nonadjacent regions of color space as, for example, some best-example responses for a "grue" term near universal green and others for the same term near universal blue.

We thank David Bimler for discussion of some of the issues treated here, Mathew Dryer for help in identifying the families of the languages listed in Table 1, and Tony Belpaeme for drawing our attention to some inaccuracies in an earlier version of the WCS data processing. This material is based on work supported by National Science Foundation Grants 0130420, 0418283, and 0418404.

1. Berlin, B. & Kay, P. (1969) *Basic Color Terms: Their Universality and Evolution* (Univ. of California Press, Berkeley).
2. Kay, P. & McDaniel, C. (1978) *Language* **54,** 610–646.
3. Kay, P. & Maffi, L. (1999) *Am. Anthropol.* **101,** 743–760.
4. Heider, E. R. (1972) *J. Exp. Psychol.* **93,** 10–20.
5. Heider, E. R. & Olivier, D. C. (1972) *Cognit. Psychol.* **3,** 337–354.
6. Roberson, D., Davies, I. & Davidoff, J. (2000) *J. Exp. Psychol. Gen.* **129,** 369–398.
7. Davidoff, J., Davies, I. & Roberson, D. (1999) *Nature* **398,** 203–204.
8. Kay, P. & Regier, T. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 9085–9089.
9. Kay, P., Berlin, B., Maffi, L. & Merrifield, W. (1997) in *Color Categories in Thought and Language*, eds. Hardin, C. L. & Maffi, L. (Cambridge Univ. Press, Cambridge, U.K.), pp. 21–56.
10. Cook, R. S., Kay, P. & Regier, T (2005) in *Handbook of Categorization in Cognitive Science*, eds. Cohen, H. & Lefebvre, C. (Elsevier, London), in press.
11. MacLaury, R. E. (1997) *Behav. Brain Sci.* **20,** 202–203.
12. Kay, P. & Berlin, B. (1997) *Behav. Brain Sci.* **20,** 196–202.
13. Yendrikhovskij, S. N. (2001) *J. Imaging Sci. Technol.* **45,** 409–417.
14. Shepard, R. N. (1992) in *The Adapted Mind*, eds. Barkow, J., Cosmides, L. & Tooby, J. (Oxford Univ. Press, Oxford), pp. 495–532.
15. Jameson, K. & D'Andrade, R. G. (1997) in *Color Categories in Thought and Language*, eds. Hardin, C. L. & Maffi, L. (Cambridge Univ. Press, Cambridge, U.K.), pp. 295–319.
16. Steels, L. & Belpaeme, T. (2005) *Behav. Brain Sci.*, in press.
17. Webster, M. A. & Kay, P. (2005) in *The Anthropology of Color*, eds. MacLaury, R. E., Paramei, G. V. & Dedrick, D. (John Benjamins, Amsterdam), in press.
18. Kay, P. & Kempton, W. (1984) *Am. Anthropol.* **86,** 65–79.
19. Witthoft, N., Winawer, J., Wu, L., Frank, M., Wade, A. & Boroditsky, L. (2003) in *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (book on CD-ROM), eds. Alterman, R. & Kirsh, D. (Lawrence Erlbaum, Mahwah, NJ).

PSYCHOLOGY