

# Cross-species differences in color categorization

JOËL FAGOT

*CNRS, INCM, and Université de la Méditerranée, Marseille, France*

and

JULIE GOLDSTEIN, JULES DAVIDOFF, and ALAN PICKERING

*Goldsmiths College, University of London, London, England*

Berlin and Kay (1969) found systematic restrictions in the color terms of the world's languages and were inclined to look to the primate visual system for their origin. Because the visual system does not provide adequate neurophysiological discontinuities to supply natural color category boundaries, and because recent evidence points to a linguistic origin (Davidoff, Davies, & Roberson, 1999), a new approach was used to investigate the controversial issue of the origin of color categories. Baboons and humans were given the same task of matching-to-sample colors that crossed the blue/green boundary. The data and consequent modeling were remarkably clear-cut. All human subjects matched our generalization probe stimuli as if to a sharp boundary close to the midpoint between their training items. Despite good color discrimination, none of the baboons showed any inclination to match to a single boundary but rather responded with two boundaries close to the training stimuli. The data give no support to the claim that color categories are explicitly instantiated in the primate color vision system.

The discontinuities that humans see in the color spectrum are so obvious that there is an inclination for cognitive psychologists to believe them inevitable and, hence, to propose that color categories are intrinsic to early levels of the neurophysiology that underpins color vision (Berlin & Kay, 1969; Franklin, Clifford, Williamson, & Davies, 2005; Kay & McDaniel, 1978; MacLaury, 1992; Pitchford & Mullen, 2002; Ratliff, 1976). However, even though color categories require discontinuities in perception, not even primary color categories (red, yellow, green, and blue) could be derived from any discontinuities in the spectral sensitivity of the three cone types by Sperling and Harwerth (1971) or, as Kay and McDaniel (1978) proposed, from the output of opponent process cells (Abramov & Gordon, 1994; Webster & Mollon, 1991). So even though some current research still stresses the role of precortical color processing in the formation of color categories, their appearance must be the result of organization at higher levels (Abramov & Gordon, 1994; Gegenfurtner & Kiper, 2003).

With respect to how color categories might be implemented in cortical visual areas, it is known that cells in V1 may be responsive to quite narrow ranges of wavelength and brightness (Yoshioka, Dow, & Vautin, 1996) with fairly much the same selectivity higher up at V4 (Schein & Desimone, 1990). Some color vision scientists (e.g.,

Hanazawa, Komatsu, & Murakami, 2000; Okajima, Robertson, & Fielder, 2002; Zeki, 1983) have proposed that such cells are the origin of color categories. There would seem to be even more reason to site color categories in the inferotemporal cortex, for lesions to that area produce achromatopsia (Cowey, Heywood, & Irving-Bell, 2001) and other (noncolor) disorders of perceptual categorical processing (Wilson & DeBauche, 1981). Yet there is no evidence for any cells, even in the inferotemporal cortex, that have the properties necessary for categorical perception proposed by Harnad (1987). The argument from categorical perception should predict, for example, "green" cells as being least sensitive to color change in the middle of the category and most sensitive to boundary colors (e.g., chartreuse or turquoise).

Despite this lack of a neurophysiological underpinning, what we shall call the nativist argument gains support from the observation, extended by Franklin and Davies (2004), that 4-month-old babies show a preference for looking at blue stimuli after habituating to green (Bornstein, Kessen, & Weiskopf, 1976). There are also a few studies with nonhuman primates that point to the nativist position with respect to color categories (Matsuzawa, 1985; Sandell, Gross, & Bornstein, 1979). There are, however, concerns about Sandell et al.'s stimuli. In their study, cross- and within-category distances were simply equated by wavelength differences; these do not take into account the spectral sensitivity of the primate eye. Indeed, in every contrast, their cross-category contrasts were perceptually easier than the within-category contrasts.

Better evidence for the nativist position comes from Matsuzawa (1985), who trained a chimpanzee to associate symbols to human prototypical colors. Although most

---

This study was supported by ESF Eurocores OMLL and CNRS OHLH Programmes, and European Community Grant SEDSU 012-984. Correspondence concerning this article should be addressed to J. Fagot, CNRS, INCM, 31 chemin Joseph Aiguier, 13402 Marseille cedex 20, France (e-mail: fagot@incm.cnrs-mrs.fr).

of the generalization to other colors could be based simply on perceptual similarity, there was an exception for green that was clearly categorical. Despite the fact that the chimpanzee was trained on a dark green, she "labeled" light greens as green rather than yellow. Nevertheless, these categories did not seem well formed in a chimpanzee with only 2 years' experience with color symbols (Matsuno, Kawai, & Matsuzawa, 2004).

Data from older children and adults has given rise to an alternative, linguistic proposal—which we shall call Whorfian—for the origin of color categories (Davidoff, 2001; Davidoff, Davies, & Roberson, 1999; Kay & Kempton, 1984; Roberson, Davidoff, Davies, & Shapiro, 2004, 2005; Whorf, 1956). Unlike in the pioneering cross-linguistic work of Rosch (Rosch Heider & Olivier, 1972), the Whorfian studies found that memory and perceived (category) similarity were predicted simply by the color terms in a speaker's language. Much of this experimental work has concentrated on the distinguishing of blue and green. Inspection of a subjectively equal color space, such as that in the Munsell system, shows that around half of the colors come from these two categories. Thus, blue and green provide a bigger range of possible within- and between-category exemplars. However, no studies have expressly tested the Whorfian hypothesis by searching for the absence of a blue/green boundary in nonhuman primates.

Old-world monkeys such as baboons are known to have wavelength discrimination functions similar to those in humans (Adams, Bryan, & Jones, 1968; Fobes & King, 1982), presumably related to the similar spectral absorption properties of their cone types (Bowmaker, Astell, Hunt, & Mollon, 1991). Also, although there is indication of somewhat different retinal distributions of short wave sensitive cones among primates (Roorda, Metha, Lennie, & Williams, 2001), synaptic connectivity, including that of baboons, appears to be very similar (Martin, Grünert, Chan, & Ghosh, 2001). Given these cross-species similarities, the nativist argument would propose the same green and blue categories for baboons as for humans; the Whorfian argument clearly would not.

In the present study, we sought to examine again the fundamental issue of whether or not there is a natural discontinuity in the human green and blue regions for nonhuman primate vision. Rather than teach monkeys the range of colors to be labeled green or blue, we simply asked whether there might be a natural boundary within that range. Thus, in the present study, we examined blue and green categories in the simplest procedure inclined to produce categorical responding (McKone, Martini, & Nakayama, 2001); we then modeled the data to verify the underlying category boundaries.

## METHOD

### Subjects

The subjects were 5 male and 3 female adult Guinea baboons (*Papio papio*) who lived in the same social group at the CNRS INCM institute. The baboons were already familiar with the matching-to-sample procedure (MTS; Fagot, Wasserman, & Young, 2001) and

were not food deprived. Five men and 3 women also served as subjects for payment. They all had normal color vision as tested by Ishihara plates (Ishihara, 1998). The adults varied in age from 21 to 37 years (mean, 25.7 years).

### Apparatus

The experiment took place in a darkened room. The baboons were tested in an experimental enclosure (60 × 50 × 72 cm) facing an analog joystick, a metal touch pad, and a 14-in. color monitor driven by a Pentium 4 personal computer. On the front of the enclosure were a viewing port, a hand port, and a food dispenser delivering banana-flavored food pellets into the enclosure. Manipulation of the joystick induced isomorphic displacements of a cursor on the monitor. The monkey viewed the screen through a viewing port, which maintained the eye–screen distance at approximately 49 cm.

For the human subjects, the same monitor and joystick were placed on a table so that the viewing distance remained approximately equal to 49 cm. Control and randomization of conditions were achieved with programs written in E-Prime 1.0 (Psychology Software Tools). Color calibration was achieved as in Roberson and Davidoff (2000).

### Stimuli

The stimuli were derived from Munsell colors (Munsell, 1905). All colors in the present studies had constant brightness (level 5) and saturation (level 6). Twelve colors were chosen to be linearly spaced (with 2.5 Munsell hue units between adjacent colors), going from 2.5G to 10B. The color stimuli subtended 6.4° × 6.4° of visual angle.

### Procedure

In each trial, the baboons placed one hand on the touch pad, which produced a .5° circular green cursor along with a white .5° × .5° square-shaped stimulus on the screen, located 1.5° above or below the cursor. The baboons had to manipulate the joystick to place the green dot on the white square to initiate the MTS procedure; then a square-shaped sample stimulus appeared with 4.5° of lateral eccentricity on the right or the left of the screen. The sample stimulus was displayed for 500 msec and was immediately replaced by a display comprising a cursor and two patches of color for comparison with the preceding sample. In balanced order, one color square appeared on the top half of the screen, and the other on the bottom half. In response to the display, baboons had to point with the cursor to the comparison stimulus that matched the sample. There was no time limit for responding. Response choices and response times were recorded. Baboons sometimes received a food pellet inside the enclosure, in accord with the prevailing reinforcement contingencies.

In training, only 2.5G and 10B stimuli were used as the sample and comparison stimuli. Training sessions comprised 96 randomly ordered and differentially reinforced identity matching trials, resulting from a completely balanced stimulus identity × stimulus position design. Subjects from both species were required to reach a criterion level of 80% correct in training trials before being tested.

Experimental test sessions comprised identity (baseline) and similarity (probe) MTS trials. In the identity trials, the target was identical to one of the comparison stimuli. In the similarity matching trials, the target was randomly chosen from the 10 intermediate colors from 5G to 7.5B. For all trials, the color comparison stimuli were 2.5G and 10B. All possible combinations of stimulus position were given equally often in a random order. The 2 stimuli used on identity matching trials were each shown 15 times, and the 10 stimuli used on similarity matching trials were each shown once, for each combination of sample and comparison stimulus position, resulting in a total of 160 trials per session. The proportion of identity and similarity matching trials (3:1) was chosen to maintain subjects' attention during the test. Ten test sessions were required for each subject and species. The baboons received a food pellet whenever they gave a correct matching response in baseline trials. Independently of re-

sponses, probe trials were randomly reinforced at an 80% rate. In cases of an incorrect baseline response, and for the remaining 20% of the probe trials, the baboons received a 3-sec time-out period with a black screen. The unique procedural differences for humans, apart from verbal instructions and their having to discover the matching rule by themselves, was that they were allowed breaks on every occasion after completion of the first three test sessions.

## RESULTS

Training performance varied from one baboon to the next. The sessions needed for the MTS rule to be learned varied from 4 to 26, but overall, an average of 10 sessions was needed for the baboons to reach criterion. One session sufficed for all human subjects to learn the rule.

Figure 1 illustrates the percentage of “green” responses for probe trials for each species. The frequency of selection of the extreme green color (i.e., 2.5G) was considered as the dependent variable in a species (human or baboon)  $\times$  probe sample color (10 possible colors, from 5G to 7.5B) ANOVA. The analysis indicated a significant main effect of probe color [ $F(9,126) = 177.9, p < .001$ ], as well as a significant probe color  $\times$  species interaction [ $F(9,126) = 34.5, p < .001$ ].

There were clearly different response profiles in baboons and humans across the 10 probe stimulus types. Curve fitting of the data using polynomial contrasts revealed a highly significant cubic trend  $\times$  species interaction [ $F(1,14) = 137.38, p < .001$ ]. Analysis of each species separately showed highly significant cubic trends for humans [ $F(1,7) = 170.4, p < .001$ ] and baboons [ $F(1,7) = 14.71, p < .006$ ], with these trends having opposite curvatures in the two species.

## MODELING

Central questions for this research were whether the subjects’ categorization behaviors could be characterized

by a “boundary” between subsets of the probe stimuli, and whether human and baboon subjects use boundaries in similar ways. Fortunately, categorization behavior has been analyzed using formal decision boundary (DB) models (Ashby, 1992) postulating the existence of such boundaries. These models have proved extremely successful in capturing human performance on a range of category learning tasks (Maddox & Ashby, 1993).

We constructed a hierarchy of simple DB models to fit to the probe trial responses made by each subject. The models assume that each probe trial stimulus creates a perceptual response (denoted  $p_i$  for the  $i$ th probe stimulus, with  $i$  ranging from 1 to 10 from the most green probe [5G] to the most blue [7.5B]). The perceptual response,  $p_i$ , across repeated presentations of the same probe stimulus, is assumed to vary due to perceptual noise;  $p_i$  is assumed to be drawn from a random normal distribution with a mean  $x_i$  and variance  $\sigma_{p_i}^2$ . The simplest DB model of the present data assumes that the means ( $x_1$ – $x_{10}$ ) are equally spaced, consistent with the equal Munsell spacing of the probe stimuli. We can arbitrarily set the mean values of  $x_1$  to  $x_{10}$  to be 1 to 10, because the values are scaled by other (noise) parameters in the model. Furthermore, the variance of the noise is assumed to be independent of the mean and the same for each probe stimulus; it is thus represented by a single parameter,  $\sigma_p^2$ .

The DB model proposes that the subject sets a simple linear boundary at a particular value within the range of perceptual response values created by the probe stimuli. The perceptual response value corresponding to this boundary is denoted by  $k$ . For each probe trial, if the resulting perceptual response is greater than  $k$ , then the subject responds by matching to the blue training stimulus (10B); if the perceptual response is less than  $k$ , then the subject responds by matching to the green training stimulus (2.5G). The same boundary is employed for all probe stimuli, but its position varies across trials with zero-mean

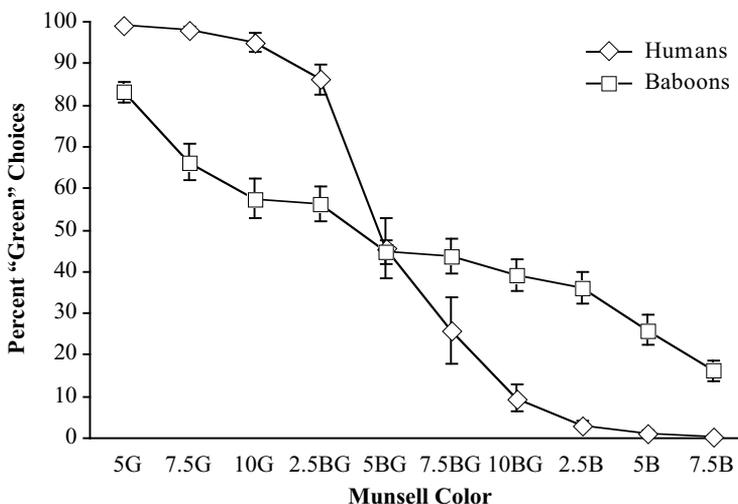


Figure 1. Response choices (percent of “green” responses) and standard errors for human and baboon subjects.

Gaussian noise, of variance  $\sigma_g^2$ . This noise can be collapsed together with the perceptual noise term ( $\sigma_p^2$ ) into a single overall noise variance parameter,  $\sigma^2$ . One estimates the probability of green responses produced for each probe stimulus from the normal distribution. The probability of a green response for probe stimulus  $i$ ,  $P(G|i)$ , is given by

$$P(G | i) = \Phi\left(\frac{k - x_i}{\sigma}\right), \tag{1}$$

where  $\Phi(z)$  is the cumulative distribution function of the standard normal distribution. The two-parameter model, given by Equation 1, was fit, for each subject individually, to the probability of green responses made to each of the 10 probe stimulus types. Fitting used a constrained iterative routine in MATLAB to obtain maximum likelihood estimates of the two parameters ( $\sigma^2$  and  $k$ ).

We also modeled the data with a more elaborate four-parameter model. We assumed that subjects adopted two boundaries, one closer to the perceptual response elicited to the green training stimulus (at perceptual response value,  $k_G$ ), and another closer to the blue training stimulus (at perceptual response value,  $k_B$ ). If a probe stimulus produces a perceptual response below  $k_G$ , then the response given is to match to the green training stimulus (2.5G); if the perceptual response is above  $k_B$ , then a match is made to the blue training stimulus (10B). A guess is produced if the perceptual response lies at or between the two boundaries. On the guessing trials, the probability of matching to 2.5G is given by a parameter,  $g$  ( $0 \leq g \leq 1$ ); the probability of matching to 10B is  $(1 - g)$ . Thus, the four-parameter model can be written formally as

$$P(G | i) = (1 - g) * \Phi\left(\frac{k_G - x_i}{\sigma}\right) + g * \Phi\left(\frac{k_B - x_i}{\sigma}\right). \tag{2}$$

The characteristic curve produced by the two-parameter single-decision-boundary model resembles the mean performance of the human subjects in Figure 1 (convex then concave, from left to right). By contrast, the curve produced by the four-parameter double-decision-boundary model resembles the mean performance of the baboons in Figure 1 (concave then convex, from left to right). It was therefore anticipated that the two-parameter model would better be able to account for human performance, whereas the four-parameter double-boundary model would better be able to account for baboon performance.

The model parameters were constrained in order to compare the two models appropriately. First, if  $g$  approaches 0 or 1, then the four-parameter model reduces to the two-parameter model (see Equations 1 and 2), so  $g$  was restricted to the range .2–.8. Second, if the noise parameters become very large under either model, then the curve tends toward linearity and the models become indistinguishable from one another. The probe stimuli were modeled by mean perceptual response values of 1–10; the noise variance parameter,  $\sigma^2$ , was constrained to the 0–10 range (almost all the good fits were achieved with values from 1 to 3; see Table 1). Last, the positions of the boundaries were restricted: for the two-parameter model, the single boundary ( $k$ ) was constrained to lie somewhere from 0 to 11 (i.e., the model values of the mean perceptual responses produced by the two training stimuli). For the four-parameter model, the green boundary ( $k_G$ ) was constrained to lie from 0 to 5.5; the blue boundary ( $k_B$ ) was constrained to lie from 5.5 to 11.

Likelihood ratio (LR) test statistics were computed in order to compare the fits of the models (for a review of this standard method for comparing model fits, see Maddox & Ashby, 1993). The two-parameter model was con-

**Table 1**  
**Best-Fitting Models and Parameter Values**  
**for Individual Human and Baboon Subjects**

Subject	Better-Fitting Model (2 or 4 Parameters)*	$k$	$k_B$	$k_G$	$g$	$\sigma^2$	LR†	$df$	$p$
Human 1	2 (.28)	4.61	–	–	–	2.30	21.3	8	.0064
Human 2	2 (.95)	5.01	–	–	–	0.74	3.84	8	.87
Human 3	2 (1.00)	5.34	–	–	–	4.27	6.34	8	.61
Human 4	2 (.08)	5.32	–	–	–	1.32	9.09	8	.33
Human 5	2 (1.00)	4.49	–	–	–	1.56	21.0	8	.0071
Human 6	2 (.45)	5.73	–	–	–	1.38	6.00	8	.65
Human 7	2 (.67)	6.05	–	–	–	2.28	21.95	8	.0050
Human 8	2 (.08)	4.69	–	–	–	1.59	20.89	8	.0074
Baboon 1	4 (.01)	–	1.21	9.64	.40	0.74	1.60	6	.95
Baboon 2	4 (.002)	–	3.20	9.37	.28	10.00	1.96	6	.92
Baboon 3	4 (.02)	–	2.09	8.14	.71	3.01	6.79	6	.34
Baboon 4	4 (<10 <sup>-14</sup> )	–	1.20	10.41	.51	3.13	10.11	6	.12
Baboon 5	4 (<10 <sup>-12</sup> )	–	1.52	10.02	.39	2.89	8.60	6	.20
Baboon 6	4 (.014)	–	2.31	8.80	.41	1.56	5.82	6	.44
Baboon 7	4 (<10 <sup>-14</sup> )	–	0.94	9.41	.47	2.13	3.38	6	.76
Baboon 8	4 (.011)	–	2.50	10.57	.32	2.99	2.90	6	.82

\*The  $p$  value given in parentheses is that associated with the likelihood ratio (LR) test comparing the two- and four-parameter models ( $p > .05$  means that the two-parameter model is preferred). †The LR statistic compares the better-fitting model with the perfect saturated model.

sidered the better-fitting model when the LR test statistic (comparing the two- and four-parameter models) was nonsignificant—that is, when its value did not exceed the 95th percentile of the  $\chi^2$  distribution (with degrees of freedom equal to the difference in numbers of parameters between the two models;  $df = 2$  in this case). As a final test of fit, the better-fitting model (with either two or four parameters) was then compared, via another LR test statistic, with a perfect saturated model. This saturated model used 10 parameters (the observed probabilities of “green” responses for each of the 10 probe stimulus types), and was (trivially) able to predict the observed probabilities perfectly. When this second LR test statistic is nonsignificant (below the 95th percentile of the  $\chi^2$  distribution, with  $df = 6$  or  $df = 8$  as appropriate), then the simpler model is not significantly worse than the perfect model, and it can be deemed an extremely well-fitting model.

Table 1 shows the model-fitting results for each baboon and human subject. Clearly, the human data were better fit by the two-parameter model than by the four-parameter model (8 out of 8 humans); the baboons, by contrast, were better fit by the four-parameter model than by the two-parameter model (8 out of 8 baboons). For 4 of the human subjects, even the saturated model was not a better fit than the two-parameter model. The baboon data, with the exception of Baboon 2, were extremely well fit by the four-parameter model. For Baboon 2, the best-fitting noise parameter was the maximum allowed (10.0), a much higher value than those in any of the other model fits. Such a high noise value tends to remove the curvature from the curve produced under the model and allows the model to fit the data artifactually. However, the fit of the two-parameter model was significantly worse than that of the four-parameter model, and the best-fitting two-parameter model was also obtained with the maximum noise parameter of 10.0. For this subject, it seems that neither model truly captured the data. For the remaining 7 baboons, the fit was excellent; in all cases, the fit of the four-parameter model was not significantly inferior to that of the saturated model (which captured the data exactly). For all of the human subjects, the single boundary was close to the midpoint of the probe stimuli used (range 4.49–6.05; midpoint, 5.5).

## DISCUSSION

A new approach was adopted to investigate the controversial issue of the origin of color categories. The data were remarkably clear-cut. Both humans and baboons performed well at the task, but the humans matched our stimuli as if with a sharp blue/green boundary. None of the baboons, despite good color discrimination, showed any inclination to match to a single boundary but rather responded as if with two boundaries that varied among individuals but that were relatively close to the training stimuli. It is true that the baboons took considerably longer than the humans to learn the training stimuli to criterion. However, it is very unlikely that more training would have allowed the baboon and human data on probe trials to

converge, because the curve fitting showed qualitatively distinct patterns of performance. As a means of simulating performance at an earlier stage of learning or task competence, we can take the best-fitting human model parameters and increase the noise for each subject. This simply linearizes the curvature; it does not produce the significant curvature in the opposite direction that was observed in the baboons' performance. Instead, it would appear that baboons, like patients who have lost color names (Goldstein, 1948; Roberson, Davidoff, & Braisby, 1999), base categorization decisions solely on close perceptual similarity (and guess when the test stimulus is not perceptually similar to either training item).

The baboons have been trained over 15 years to categorize many types of stimuli using a similar MTS procedure. So, why is the task with colors so different for them? Other tasks allow the baboons to categorize, with sufficient training, on the basis of small perceptual differences (see, e.g., Vauclair & Fagot, 1996), but this approach would be ineffective with the continuous gradation of colors used here, or indeed, with any other stimulus continuum. At some point on the continuum, a boundary has to be drawn and the positioning of that boundary will be arbitrary without neurophysiological or nonperceptual mechanisms to aid categorization. Indeed, it can be argued (Dummett, 1975; Roberson et al., 1999) that the task is impossible without a nonperceptual procedure to aid in the boundary decision; this is shown by the modeling data, where baboons guessed during probe testing except for stimuli close to the training items. In addition, the model, with its assumptions, is able to predict the individual baboon's behavior to an extent that is statistically indistinguishable from a perfect-fitting model (see Table 1, rightmost column; all  $ps > .05$  and mostly  $\geq .05$ ). So, on grounds of parsimony one would probably want to accept our model rather than propose additional possible factors affecting the perceptual processing of the probe stimuli differentially in the two species. Humans, by contrast, matched a probe stimulus further away from the training items in a systematic way (i.e., without guessing), even though this item was perceived to be distinct from the training item to which it was matched. Such behavior is consistent with the use of linguistic labels to aid in categorization.

The data therefore give no support whatsoever to the nativist claim that the same monkey and human color categories are explicitly instantiated in the primate color vision system. Of course, it is always possible that our design produced in baboons conditions that allowed them to ignore their innate boundary. It is admitted that our rejection of the nativist claim is based on data from only one of the many possible equally spaced sets of stimuli that go from green to blue and without manipulation of reward contingencies. However, it is entirely unclear what kind of differences between baboon and human perception of our stimuli or what change in reward contingencies would produce the precise shape of function shown by the baboons. Of course, it is possible that baboons, like chimpanzees, could be taught color categories by associating colors to a sign (Matsuzawa, 1985), but that is a different matter.

## REFERENCES

- ABRAMOV, I., & GORDON, J. (1994). Color appearance: On seeing red—or yellow, or green, or blue. *Annual Review of Psychology*, **45**, 451-485.
- ADAMS, C. K., BRYAN, A. H., & JONES, A. E. (1968). Electroretinographic determination of the spectral sensitivity of the baboon (*Papio anubis*). *Vision Research*, **8**, 1399-1405.
- ASHBY, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-483). Hillsdale, NJ: Erlbaum.
- BERLIN, B., & KAY, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- BORNSTEIN, M. H., KESSEN, W., & WEISKOPF, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 115-129.
- BOWMAKER, J. K., ASTELL, S., HUNT, D. M., & MOLLON, J. D. (1991). Photosensitive and photostable pigments in the retinae of Old World monkeys. *Journal of Experimental Biology*, **156**, 1-19.
- COWEY, A., HEYWOOD, C. A., & IRVING-BELL, L. (2001). The regional cortical basis of achromatopsia: A study on macaque monkeys and an achromatopsic patient. *European Journal of Neuroscience*, **14**, 1555-1566.
- DAVIDOFF, J. (2001). Language and perceptual categorisation. *Trends in Cognitive Sciences*, **5**, 382-387.
- DAVIDOFF, J., DAVIES, I., & ROBERSON, D. (1999). Colour categories in a stone-age tribe. *Nature*, **398**, 203-204.
- DUMMETT, M. (1975). Wang's paradox. *Synthese*, **30**, 301-324.
- FAGOT, J., WASSERMAN, E. A., & YOUNG, M. E. (2001). Discriminating the relation between relations: The role of entropy in abstract conceptualization by baboons (*Papio papio*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, **27**, 316-328.
- FOBES, J. L., & KING, J. E. (1982). *Primate behavior*. New York: Academic Press.
- FRANKLIN, A., CLIFFORD, A., WILLIAMSON, E., & DAVIES, I. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of Experimental Child Psychology*, **90**, 114-141.
- FRANKLIN, A., & DAVIES, I. (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology*, **22**, 349-377.
- GEGENFURTNER, K. R., & KIPER, D. C. (2003). Color vision. *Annual Review of Neuroscience*, **26**, 181-206.
- GOLDSTEIN, K. (1948). *Language and language disturbance*. New York: Grune & Stratton.
- HANAZAWA, A., KOMATSU, H., & MURAKAMI, I. (2000). Neural selectivity for hue and saturation of colour in the primary visual cortex of the monkey. *European Journal of Neuroscience*, **12**, 1753-1763.
- HARNAD, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1-28). Cambridge: Cambridge University Press.
- ISHIHARA, S. (1998). *Ishihara's tests for color blindness*. London: Hodder & Stoughton.
- KAY, P., & KEMPTON, W. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist*, **86**, 65-79.
- KAY, P., & McDANIEL, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, **54**, 610-646.
- MACLAURY, R. E. (1992). From brightness to hue: An explanatory model of color-category evolution. *Current Anthropology*, **33**, 137-186.
- MADDOX, W. T., & ASHBY, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.
- MARTIN, P. R., GRÜNERT, U., CHAN, T. L., & GHOSH, K. K. (2001). Retinal pathways for colour vision: Studies of short-wavelength sensitive ("blue") cones and their connections in primate retina. *Color Research & Application*, **26**(Suppl.), S112-S117.
- MATSUNO, T., KAWAI, N., & MATSUZAWA, T. (2004). Color classification by chimpanzees (*Pan troglodytes*) in a matching-to-sample task. *Behavioural Brain Research*, **148**, 157-165.
- MATSUZAWA, T. (1985). Color naming and classification in a chimpanzee. *Journal of Human Evolution*, **14**, 283-291.
- MCKONE, E., MARTINI, P., & NAKAYAMA, K. (2001). Categorical perception of face identity in noise isolates configural processing. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 573-599.
- MUNSELL, A. H. (1905). *A color notation*. Boston: Ellis.
- OKAJIMA, K., ROBERTSON, A. R., & FIELDER, G. H. (2002). A quantitative network model for color categorization. *Color Research & Application*, **27**, 225-232.
- PITCHFORD, N. J., & MULLEN, K. T. (2002). Is the acquisition of basic-colour terms in young children constrained? *Perception*, **31**, 1349-1370.
- RATLIFF, F. (1976). On the psychophysical basis of universal color terms. *Proceedings of the American Philosophical Society*, **120**, 311-330.
- ROBERSON, D., & DAVIDOFF, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, **28**, 977-986.
- ROBERSON, D., DAVIDOFF, J., & BRAISBY, N. (1999). Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition*, **71**, 1-42.
- ROBERSON, D., DAVIDOFF, J., DAVIES, I., & SHAPIRO, L. R. (2004). The development of color categories in two languages: A longitudinal study. *Journal of Experimental Psychology: General*, **133**, 554-571.
- ROBERSON, D., DAVIDOFF, J., DAVIES, I., & SHAPIRO, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, **50**, 378-411.
- ROORDA, A., METHA, A. B., LENNIE, P., & WILLIAMS, D. R. (2001). Packing arrangements of the three cone classes in the primate retina. *Vision Research*, **41**, 1291-1306.
- ROSCH HEIDER, E., & OLIVIER, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, **3**, 337-354.
- SANDELL, J. H., GROSS, C. G., & BORNSTEIN, M. H. (1979). Color categories in macaques. *Journal of Comparative & Physiological Psychology*, **93**, 626-635.
- SCHEIN, S. J., & DESIMONE, R. (1990). Spectral properties of V4 neurons in the macaque. *Journal of Neuroscience*, **10**, 3369-3389.
- SPELTING, H. G., & HARWERTH, R. S. (1971). Red-green cone interactions in the increment-threshold spectral sensitivity of primates. *Science*, **172**, 180-184.
- VAUCLAIR, J., & FAGOT, J. (1996). Categorization of alphanumeric characters by baboons (*Papio papio*): Within and between class stimulus discrimination. *Current Psychology of Cognition*, **15**, 449-462.
- WEBSTER, M. A., & MOLLON, J. D. (1991). Changes in colour appearance following post-receptor adaptation. *Nature*, **349**, 235-238.
- WHORF, B. L. (1956). Language, mind and reality. In J. B. Carroll (Ed.), *Language, thought and reality* (pp. 246-270). Cambridge, MA: MIT Press.
- WILSON, M., & DEBAUCHE, B. A. (1981). Inferotemporal cortex and categorical perception of visual stimuli by monkeys. *Neuropsychologia*, **19**, 29-41.
- YOSHIOKA, T., DOW, B. M., & VAUTIN, R. G. (1996). Neuronal mechanisms of colour categorization in areas V1, V2 and V4 of macaque monkey visual cortex. *Behavioural Brain Research*, **76**, 51-70.
- ZEKI, S. (1983). Colour coding in the cerebral cortex: The responses of wavelength-selective and colour-coded cells in monkey visual cortex to changes in wavelength composition. *Neuroscience*, **9**, 767-781.