COGNITIVE PSYCHOLOGY 3, 337–354 (1972)

The Structure of the Color Space in Naming and Memory for Two Languages[1]

ELEANOR ROSCH HEIDER[2]
Brown University

AND

DONALD C. OLIVIER
Harvard University

Ss from two cultures with markedly different color terminologies were tested on two color-judgment tasks. One was a nonverbal task of color matching from memory, while the other was a verbal task of color-naming. Both tasks were performed by 41 American Ss and 40 New Guinea Dani (who have a basically two-term color language). Multidimensional scaling based upon the four resulting sets of data yielded structures that were more similar under the memory condition than under the naming condition. For neither culture were equally distant colors confused in memory more within than across name boundaries. Retention of color images in short-term memory appears to be unaffected by wide cultural differences in the semantic reference of color words.

The "Whorfian hypothesis" has come to be the general label for a set of loosely defined points of view to the effect that reality is perceived and understood differently in different linguistic communities, and that these differences are caused, in some sense, by the language—particularly by the structure ("organization," "classification") laid upon reality by the language (cf. Whorf, 1956). Any empirical "test" of the Whorfian hypothesis is actually a redefinition of the general position in particular

[1] This research was supported in part by the Cross-cultural Project in Ethnocentrism, administered by Robert A. LeVine and Donald T. Campbell, and in part by Grant 02908-03 from the National Institute of Child Health and Development.

[2] Now at the University of California, Berkeley. We are grateful to Roger Brown for his support and ideas, Philip Dale for his suggestions about "distance control," Robert Krauss for his initial discussions, Roger Shepard for many helpful comments on the manuscript, the Cambridge Friends School for the participation of its students, and to the government of the Republic of Indonesia, Karl G. Heider, and Father Frans Verheijen, OFM, for assistance in West Irian (West New Guinea).

operational terms and a test of that specific redefinition. Whorf himself concentrated on the interpretive description of how languages actually differ. Psychological "language-cognition" studies have attempted to relate linguistic differences to specific nonlinguistic behaviors which could serve as an independent measure of perception or thought.

The major group of psychological studies which have succeeded in showing a relation between language and cognition are those which have used color as the nonlinguistic domain, memory as the cognitive variable, and semantic aspects of language as the linguistic variable. Two semantic measures have received the focus of attention. Brown and Lenneberg (1954) found that the codability of colors (a composite measure of agreement in naming, length of name, and response latency of naming) was positively correlated with accuracy of recognition memory; however, the relation could not be replicated for all arrays of color (Burnham & Clark, 1955; Lenneberg, 1961). Lantz and Stefflre (1964) showed that communication accuracy (the accuracy with which a subject's verbal description of a color allows other native speakers of the language to pick out that particular color) was positively correlated with recognition—a finding which has been replicated not only for diverse arrays of color (Lantz & Stefflre, 1964) but also cross-culturally for different languages (Stefflre, Castillo Vales, & Morley, 1966) and for stimuli other than color (Koen, 1966).

The present research did not seek to dispute these earlier findings. Our aim was, rather, to test a different kind of operational claim derived from the Whorfian position, a claim that verbal color coding acts on memory imagery such that the "structure" of colors in memory comes to resemble the "structure" of color names in a given language. By "structure," we mean perceived distance between stimuli and the resultant grouping or dimensionality of the stimuli. In order to demonstrate a relation between the structure of colors in memory imagery and the structure of color naming in natural languages, we must have: measures of naming and memory which measure structure; a task for which it is meaningful to talk about the influence of naming on memory mechanisms; and at least two natural languages whose color name categories structure the color space in different ways.

Previous studies have not employed measures of structure. As Stefflre *et al.* (1966) point out, the relation between communication accuracy and memory accuracy is a relation between inter- and intrapersonal communication. The fact that those two variables are consistently correlated does not provide information about the distance relationships of colors to each other in either communication or memory. Codability, likewise, does not measure distances among colors. The present study used two techniques to compare the structure imposed on the color space by a language's system of color categories and the pattern of errors speakers of that language made in color memory: (a) The structure of the color space inferred from multidimensional scaling of the "confusions" of color stimuli in a naming task was compared with the structure obtained from confusions in memory; and (b) A comparison was made of the extent of confusions in memory between equally "distant" colors given the same and different names.

Posner (1969) has argued for the existence of a purely visual code in memory and has shown (Posner, 1966) that memory for one kind of "purely" visual item could be impaired by interpolated tasks in the same way that rehearsal of verbal items can be impaired. Choice of a memory task in the present study involved the basic decision of whether or not to attempt (by the use of interpolated tasks) to eliminate Ss' ability to employ a visual code as a memory aid. Stefflre *et al.* (1966) asked Ss to recognize single colors after a 30-sec unfilled interval. Memory accuracy correlated with communication accuracy in spite of the fact that, in such a task, visual coding is readily available to Ss. The present study chose to use a similar task for two reasons: (a) Its comparability to the Stefflre *et al.* (1966) task meant that we could ask our structural question of a context for which one kind of relation between language and memory had already been demonstrated. (b) Most memory for visual items appear to contain visual as well as verbal components—certainly it is very difficult to argue in any experiment that Ss are completely prevented from using one or another of the components. We, therefore, used a task in which visual and verbal coding could occur simultaneously but which was sufficiently difficult such that only partially accurate memory was obtained—namely, recognition of single colors from an array of highly similar stimuli after a 30-sec unfilled interval. Our question was whether the verbal code would interact with the visual to influence the nature of memory errors.

Correspondence between the structure of color naming and the structure of memory errors in a single language would not demonstrate that the naming structure was in any sense prior to the memory—in fact, from such evidence, an argument for the reverse relation could equally be made. What was needed was a comparison of two languages with radically different structures of color categorization. A recent study (Berlin & Kay, 1969) suggested that the languages hitherto compared in psychological color research may have virtually identical structures of color terminology. Berlin and Kay argued that basic color terms were universally limited to no more than eight chromatic terms and that the focal points (best examples) of these terms were universal. Thus the differences between languages possessing a full complement of color terms (such as English, Spanish or Hopi) would lie primarily in the placement of boundaries between categories, and only languages which differed markedly from these in the number of basic color terms could be expected to divide the color solid in radically different ways.

The Dani of Indonesian New Guinea possess such a color terminology. This stone age, agricultural people had been previously described by K. G. Heider (1970) as having only two color terms, which divide the color space on the basis of brightness rather than hue. Such a color system has been reported for other cultures as well and is referred to by Berlin and Kay as Stage I, the simplest form of color terminology, allegedly entirely brightness based. In Berlin and Kay's systematization, the two terms are supposed to correspond to English "dark" and "light" with their focal points in black and white, respectively. This terminology appears maximally different from English and offers the possibility that the Dani would name and confuse colors on different dimensions altogether than would English speakers.

Our operational translation of the Whorfian hypothesis predicts that in a color memory task in which both visual and verbal coding could occur: (a) The multidimensional structure derived from memory will be the same as that derived from naming within each culture but will be different across cultures, and (b) equally "distant" colors given the same name will have higher confusion scores than those given different names.

METHODS

Subjects

The Ss were 40 Dani and 41 Americans—equal numbers of adults and children (age 8–11), males and females. The same Ss performed naming and memory. All Ss were screened for color blindness and were paid volunteers. All Dani were monolingual; American Ss were native English speakers.

Stimulus Materials

Munsell color chips of glossy finish were used and are here referred to in Munsell notation. C. I. E. tristimulus values and chromaticity coordinates are obtainable from Nickerson, Tomaszewski and Boyd (1953). The term "brightness" is here used to refer to the Munsell dimension "value."

The array used for naming and memory consisted of four brightnesses (9/, 7/, 5/, 3/) of each of ten hues evenly spaced around the spectrum

(7.5R, 7.5YR, 7.5Y, 7.5GY, 7.5G, 7.5BG, 7.5B, 7.5PB, 7.5P, 7.5RP), all at the same low saturation (/2). Chips were mounted in an array reflecting their "Munsell order"; hues horizontally, brightnesses vertically. The chips were mounted, uncovered, on white cardboard by means of insertion of their tabs into slits in the cardboard. The Farnsworth–Munsell colors (Farnsworth, 1949) were not used in this study because brightness is a dimension of particular interest in the Dani–English comparison, and the Farnsworth–Munsell colors do not provide variations in brightness. An array of more saturated colors was not used because it is in saturated colors that universal "focal areas" for color names are found which could have confounded results in the present study as was the case in Brown and Lenneberg (1954)—see Heider (1972) for a more extended treatment of this problem.

Munsell chips are not perfectly perceptually equidistant, and in any array constructed *ad hoc* such inequalities may influence results. To test for such influences in the present array, two checks were performed: (a) Ten American Ss were asked to order each column according to brightness and each row according to hue. The result was that, as with the Farnsworth–Munsell array, misorderings of hues were evenly spaced around the spectrum. Brightness ordering was highly accurate. (b) The true perceptual distance of the chips from each other was determined by use of the tables in Newhall, Nickerson and Judd (1943). The discriminability of each chip from its neighboring chips was computed using the procedure described in Brown and Lenneberg (1954) with two modifications: each chip was considered to have four rather than two neighbors on each of the three dimensions (neighbors for edge colors were obtained by extrapolation), and the Brown and Lenneberg correction for margins was not applied (there was no evidence that edge colors had a lower chance of recognition in the present array). Disparities among perceptual distances between chips of the present array were considerably smaller than those of the Brown and Lenneberg array, and discriminability differences between chips were small. These two checks indicated that, for many purposes, each hue step in the 40 chip array could be considered equivalent to each other hue step and each brightness step equivalent to each other brightness step, although the brightness steps might be larger than the hue steps.

The colors of this 40 chip array, because of their low saturation, did not represent a full range of natural colors (i.e., colors obviously visible in the Dani and American environment), nor were they equivalent to any of the colors in Berlin and Kay's (1969) array. In order to ascertain that Dani and American Ss named the 40 color array in a manner similar to their naming of saturated colors, an additional array was presented for naming only. This array consisted of every other hue of the Berlin and Kay array. Chips were of eight brightnesses (9/, 8/, 7/, 6/, 5/, 4/, 3/, 2/) of each of 20 hues evenly spaced around the spectrum (5R, 10R, 5YR, 10YR, 5Y, 10Y, 5GY, 10GY, 5G, 10G, 5BG, 10BG, 5B, 10B, 5PB, 10PB, 5P, 10P, 5RP, 10RP); each was at maximum saturation attainable for that hue and brightness (as printed in the Munsell Book of Color: glossy finish collection, 1966). Chips were mounted in the same manner as in the 40 color array.

Illumination

All tasks in both cultures were performed indoors by the illumination of natural daylight coming through a door or window. The spectral characteristics of equatorial light differ from Northern Hemisphere light; however, two facts argue that this is not a consideration in the interpretation of the present study. In regard to the naming task: Kay[3] reported similar results for the Berlin and Kay tasks under differing illuminations, including fluorescent light and direct sunlight. In regard to memory: differences in illumination between the two cultures could only cause a bias toward differences in memory, and such a bias is opposite in direction to the results actually obtained.

Procedures

Order of tasks. The two color naming tasks and the color memory task were performed by the same Ss. Order of presentation for half the Ss in each category was: naming the 40 color array, recognition memory, naming the 160 color array. For the other half, order of presentation was: naming the 160 color array, naming the 40 color array, recognition memory. At the end of testing, American Ss were asked to tell of each chip in the 160 color array, whether they would call it "light" or "dark."

Naming. For the Dani, naming instructions were (in Dani), "This here (*E* points to a chip) What is it called?" The *E* pointed to chips in random order. Dani descriptions of the chips were transcribed in their entirety.

For American Ss, naming instructions were to "give a name to each chip," to "tell what you would call it." Each designation was recorded in its entirety as spoken by the *S* with the following exception: six American Ss, after naming each chip, persisted in adding strings of varying numbers of "very"s, "dark(er)"s, etc. which rendered each of the encodings unique. These (grammatically separate) additions by

[3] P. Kay, personal communication, October 8, 1969.

those Ss were the only utterances omitted in analysis of the naming data.

Instructions to the American Ss for the designation of the chips as "dark" or "light" were essentially the same as those in the naming task except that each S named chips in an order of his own choice.

Memory. The S was shown a single test chip for 5 sec, waited 30 sec, and then was shown the entire 40 color array and asked to select from it the chip he had seen. All 40 of the chips were presented to each S in random order; each S received a different random order. During testing, the S was given no feedback concerning the correctness of his answers; thus, response to the two dimensions of the array could not have been "artificially taught" in the course of testing. The 30-sec delay was not filled by any activity. Test chips were shown for the 5-sec viewing against a background of white cardboard, and the array was covered with white cardboard during the presentation of the chip and the wait.

Multidimensional Scaling Technique

Separate 40×40 similarity matrices were constructed for the memory and for the naming data for each culture. In the memory matrix the (i,j) entry was the number of times chip i as a stimulus elicited chip j as a response. In the naming matrix the (i,j) entry was the number of Ss who gave chip i and chip j the same name; an S was counted whenever his name for chip i was identical word for word to his name for chip j. Kruskal's MDSCAL version of Shepard–Kruskal multidimensional scaling (Kruskal, 1964a, 1964b; Shepard, 1962) was applied to these matrices to give four scalings of the 40 chips in three dimensions; Euclidean distance and Kruskal's "primary approach" for tied ranks was used.

RESULTS

The facts of Dani color naming were roughly as anthropological investigation had previously claimed (K. G. Heider, 1970). There were only two color terms ("mili" and "mola") which were used by all Ss. These terms were not, however, based purely on brightness. Table 1 shows the distribution of "mili" over both arrays: "mola" was essentially its complement. Roughly speaking, "mili" included both dark and "cold" colors; "mola" light and "warm" colors. The form of the division into "mili" and "mola" was identical for the 40 color, unsaturated, and the 160 color, saturated, arrays. There was no equivalent English categorization of the spectrum—for American Ss, the terms "dark" and "light" were purely brightness based. For the 40 color array, 80% of all the names given were "mili" or "mola." The remaining 20% were not terms used

TABLE 1
Distribution of the Dani Color Name "Mili" (Roughly "Dark")[a]

The 40 color array

Brightness	7.5R	7.5YR	7.5Y	7.5GY	7.5G	7.5BG	7.5B	7.5PB	7.5P	7.5RP
9/	2	2	3	3	6	8	5	4	1	1
7/	5	6	12	20	23	26	19	16	8	5
5/	13	20	26	29	31	33	33	26	21	19
3/	25	23	29	35	34	33	32	32	23	21

The 160 color array

Brightness	5R	10R	5YR	10YR	5Y	10Y	5GY	10GY	5G	10G	5BG	10BG	5B	10B	5PB	10PB	5P	10P	5RP	10RP
9/	0	0	3	3	5	2	3	3	5	5	5	4	1	3	2	2	1	2	1	0
8/	0	1	2	2	3	4	8	18	19	21	13	15	13	10	11	7	5	2	0	1
7/	1	1	3	4	7	7	15	23	25	26	29	22	21	20	17	14	8	2	0	0
6/	0	0	2	7	10	16	20	23	31	31	29	29	26	26	22	19	7	2	0	0
5/	0	0	1	11	16	17	24	26	35	30	28	26	26	27	23	19	6	0	0	0
4/	1	2	10	18	19	28	24	30	30	31	32	29	29	26	30	18	14	4	2	0
3/	2	9	24	22	30	33	36	34	33	36	37	35	34	36	30	24	10	8	3	3
2/	12	23	26	30	35	37	37	37	38	39	37	35	35	34	30	25	14	9	11	7

[a] Note. Numbers represent the number of Dani *S*s who named the designated chip "mili."

to refer consistently to the same unsaturated colors by any portion of the subject population; these terms were probably idiosyncratic descriptions rather than basic color names. All data analysis was based on the full name given by each S.

The distribution of American basic color names held few surprises. The terms were primarily hue based, although "red," "pink," "yellow," "orange," and "brown" had a brightness component. A simplified picture of the distribution of American basic hue terms over the Munsell chips (the chips shown for Dani terms in Table 1) is available in Berlin and Kay (1969). Almost all of the American Ss used additional nonbasic names, including less common English color names ("turquoise," "peach"), combinations of basic names ("blue-green"), and elaborated idiosyncratic descriptions ("sick-maroon," "British racing car green"). These nonbasic names accounted for about half of all names given for the 40 color array. As with Dani names, American data analysis was based on the full name given by each S.

On the memory tasks, the Dani Ss performed much less accurately than the American Ss. The mean number of correct identifications for Dani was 7.7, with a range of 0–17; for Americans, the mean was 11.7, with a range of 4–28. The difference was highly significant: $t = 8.1$, $p < .001$.

Scaling Results

The scaled configurations were compared both impressionistically and also using a "symmetric measure of fit" for nonmetric scaling solutions devised by Schönemann and Carroll (1970). Neither method by itself is entirely satisfactory; visual examination of the scalings is only suggestive, but the measure of fit, which gives a numerical measure of the difference between two configurations, loses most of the useful detail of the scaling solutions. For want of a better technique, we present both comparisons.

Preliminary examination of the scalings indicated that all four configurations were roughly cylindrical, as if the rectangular stimulus array was "wrapped around" the brightness axis to bring the red edge near the purple edge. As an aid to visualization, the scalings were rotated to bring them as close as possible into the same orientation with the brightness axis along the first (vertical) dimension, using the "orthogonal Procrustes technique" described in Schönemann (1966) and Cliff (1966). The four rotated scalings are shown in Figs. 1 and 2; for compactness of presentation, only the first two dimensions are plotted, as if the color cylinders were viewed from the side. For the memory scalings, chips of equal brightness are connected by lines to organize the picture.

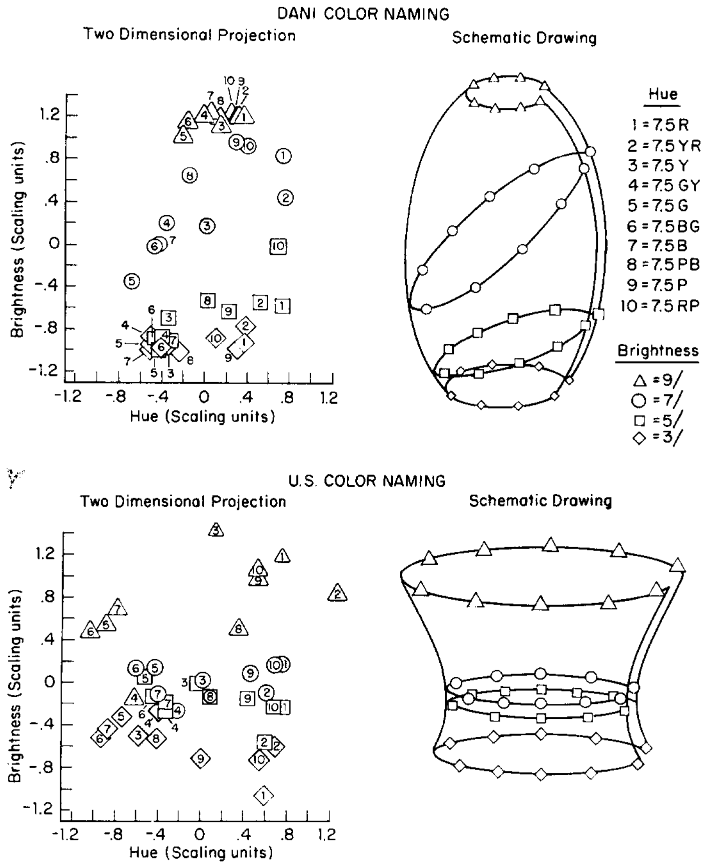

FIG. 1. Multidimensional scaling of Dani and American color naming.

Schematic drawings of the three-dimensional scaling configurations are provided beside the two-dimensional plots as an aid to interpretation. (The dimension labeled "hue" in the figures corresponds to the Munsell dimension labeled "hue," roughly to wavelength of the reflected light; "perceived hue"—as demonstrated by Shepard, 1964, and by Garner and Felfoldy, 1970—may include other psychologically "integral" dimensions, such as intensity and saturation.)

Apart from their roughly cylindrical form, the naming scalings for the two cultures look different, as one might expect from the differences in Dani and American color terminology, but the memory scalings do not show corresponding differences. In the Dani naming scaling, the brightest chips and the darkest chips fell into two small clusters, as if the color cylinder were constricted at each end. These chips were almost

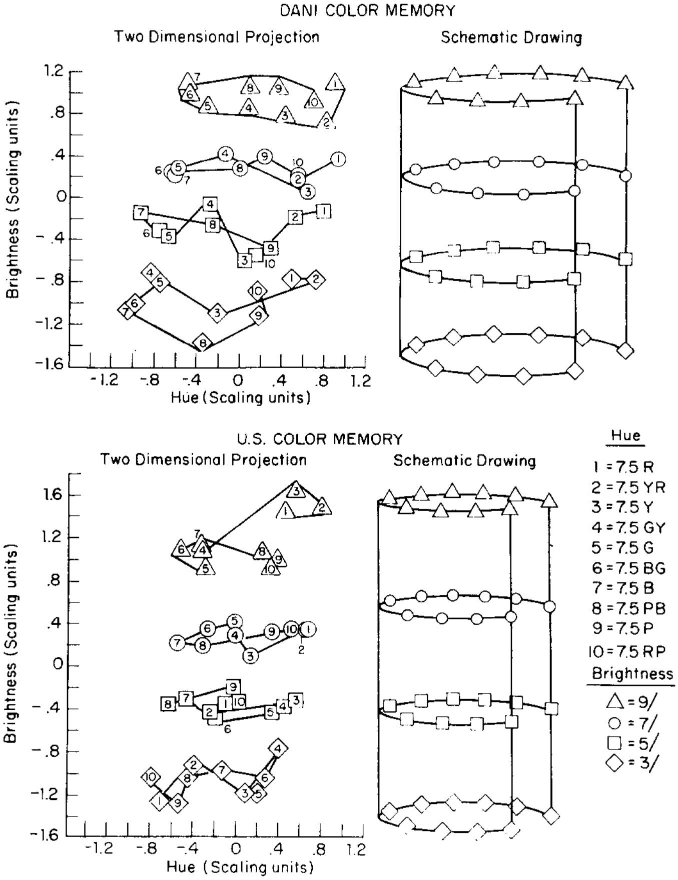

FIG. 2. Multidimensional scaling of Dani and American color memory.

always called "mola" and "mili," respectively, and, hence, were very little differentiated by hue. Chips of intermediate brightness were more widely separated, primarily because Ss differed in their placement of the boundary between the two basic terms (see Table 1). Red and purple chips—those more likely to be called "mola"—were closer to the upper end of the brightness axis than blue and green chips. In the American naming scaling, hue distinctions were more salient than brightness distinctions, and, in contrast to the Dani scaling, chips of extreme brightness were more widely separated than chips of intermediate brightness; this reflected the fact that a group of names including "white," "yellow," "pink," and "black" were used to distinguish among chips of extreme brightness but were seldom used for chips of intermediate brightness.

In the memory scalings for both cultures, the configurations were more nearly cylindrical, with chips of the four brightnesses spaced almost equally along the brightness axis. There are no visible features in the memory scalings corresponding to the construction at extreme brightness in the Dani naming scaling, the spreading at extreme brightness in the American naming scaling, or the skewing of red chips toward the bright end of the cylinder in the Dani naming scaling. The only clear difference between the two memory scalings is the relative elongation of the American scaling along the brightness axis. Taken at face value, this would imply that American Ss had better memory for brightness, relative to hue, than Dani Ss, but it might simply be an effect of the superior overall accuracy of the American Ss.

The values obtained for Schönemann and Carroll's measure of departure from good fit were as follows: Dani vs American naming—.194; Dani vs American memory—.161; Dani naming vs memory—.126; American naming vs memory—.212. Smaller numbers correspond to better fit, with a theoretically possible range for data like ours[4] from a minimum value of 0 for identical configurations to a maximum of .667 for completely unrelated configurations. The quasi-cylindrical nature of all of our configurations further constrained the possible dissimilarity between the four scaling solutions; thus, the maximum score for unrelatedness which our configurations could be expected to achieve was probably in the approximate range .250–.300. The Whorfian hypothesis would lead one to expect that the second two comparisons, across modalities but within cultures, would yield better fits and hence smaller numbers than the first two, across cultures but within modalities. The data did not confirm this prediction. (For example, if Dani naming vs memory and American naming vs memory had both received measures of fit $< .100$, and Dani vs American naming and Dani vs American memory had both received measures $> .200$, it would have constituted evidence for a Whorfian view; however, the actual data differed from such a pattern.) Visual impression of the configurations (in addition to not confirming the Whorfian prediction) also suggested that the memory configurations for the two cultures were more similar than were any of the other comparisons. The numerical measures of fit supported the visual impression that the Dani and American memory configurations were more similar to each other than were the two cultures' naming configurations. The measures were not in perfect accord with subjective judgments, however, in that they gave a better fit between Dani naming

[4] That is, for scalings in three dimensions with the scale factor chosen so that the mean distance of points from the origin is 1.

and memory than between the two memory scalings. It is not clear how to account for the discrepancy; it may be that the eye discounts the difference in overall proportions—length vs width—of the two memory scalings, while Schönemann and Carroll's measure does not.

In summary: Although the numerical measures of similarity between the four scalings were not unequivocal, neither they nor the shapes of the configurations themselves gave any evidence that differences between the color name structures of the two cultures were reflected in equivalent differences in the patterns of confusions in the memory task.

Confusions of Chips Given the Same Name

There was a good deal of variation among Ss in each culture, both in naming and in memory task performance, and it was possible that there were relationships between naming and memory for each individual which were obscured when the results were pooled. The most straightforward hypothesis is that an individual will be more likely to confuse colors to which he gave the same name than those to which he gave different names. Previous research has not tested this hypothesis directly because of the confounding factor of distance; colors given the same name and colors confused in memory are the colors closest to each other in an array. Distance was controlled in the present analysis by considering only pairs of colors adjacent in the array—chips of the same hue and adjacent brightness, or same brightness and adjacent hue. Table 2 shows the analysis of accuracy for adjacent colors with the same and with different names. Three of the four comparisons were in the

TABLE 2
Memory Confusions of Adjacent Colors[a]

Naming condition	Percentage of confusions: U. S.	Percentage of confusions: Dani
	Adjacent hues	
Same name	26.8	16.9
Different name	25.0	14.5
	Adjacent brightnesses	
Same name	12.4	12.3
Different name	9.5	13.2

[a] Note. χ^2 test performed for each same name-different name pair. None of the differences were significant.

right direction, but all the differences were small and none reached significance by the chi square test. Here again there was no evidence of a significant effect of naming on memory.

DISCUSSION

The structure of the color space derived from the color naming data of American English speaking Ss and of the Dani of Indonesian New Guinea were similar to the extent that they were both quasi-cylindrical but otherwise quite different. A "structural" interpretation of the Whorfian hypothesis would assert that those differences should be reflected in equivalent cultural differences in the structure of the color space in memory. We derived the structure of the same space in memory from color confusions in the memory task. There was no indication that the differences between the naming structures for the two languages carried over in parallel fashion to the two memory structures. Furthermore, in a comparison of memory confusions for equally "distant" pairs of colors which had been given the same versus different names, there was no evidence that colors were confused more within than across naming boundaries. Not proven, but certainly suggested by the visual shape of the scaling configurations, was the further finding that Dani and American color memory structures were quite similar to each other, although the naming structures were not. These results do not contradict previous studies, which dealt with a different operational definition of the Whorfian hypothesis; however, the difference requires explanation.

The color memory task—both ours and Stefflre *et al.* (1966)—required the *S* to recognize a color, after a 30-sec time interval, from among other colors very similar to the original. To do this, the S had to "retain" a single color in memory for 30 sec; it is the mechanism of this retention which is of interest. The time period is far too long for "iconic" memory to be operative (cf. Neisser, 1967). Single auditory-verbal-linguistic items can be maintained without error in memory for 30 sec by the process of active rehearsal. Posner (1966; 1969) presented evidence for the retention of a purely visual code in short-term memory. Our task was specifically chosen to be one in which Ss could simultaneously use both types of code.

That both visual and verbal memory codes were, in fact, employed by Ss in the experimental task used is argued both by the data and by introspective reports. Visual storage is strongly suggested by the fact that color memory did not show language effects, yet was also not random; while verbal coding often occurred overtly—both American and Dani Ss, at times, named the color aloud during the memory task. As for introspective reports, American Ss, at the termination of the memory task, generally designated both strategies—i.e., thinking of a name or verbal description of the color while, simultaneously, actively trying to maintain a "mental image" of the color. The few Dani Ss with whom the *E* proved capable of communicating about such questions gave similar reports. Within this framework, we can now specify the effects that previous studies have shown linguistic coding to have on color memory and the important effect which the present study has shown it does not have on memory.

The correlation between communication accuracy and memory accuracy (Lantz & Stefflre, 1964) is a result, we propose, of the fact that verbal codes for some colors are specific enough to bypass visual storage entirely. That is, presumably each culture has a variety of secondary color terms which designate relatively limited areas of the color space or which name a well-known cultural object of a specific color (such as "olive drab"). If a code is so specific to a color in a given array that it enables a listener who has not seen the color to identify the color correctly, then similarly a *S* in a memory experiment who retains only the verbal code in short-term memory can recognize the color on the basis of the code alone just as he would if he were the listener in a communication experiment.

It is, in fact, tempting to attribute the generally poorer Dani memory performance to a lack of specific color codes such as these. However, as discussed in Heider (1972), there are many other differences between Dani and American culture to which overall memory performance could equally be attributed. Dani life contains neither school, work, nor interpersonal relations which appear to produce overloads of information; thus, Dani appear to have neither need for, practice with, nor any explicit training in the use of memory control processes. An "interaction" between culture and the type of memory errors which occurred might have demonstrated a specific effect of language on memory; a simple difference between cultures in the direction of Dani memory inferior to American does not constitute such a demonstration.

Presumably Ss sometimes also remember a color correctly on the basis of the visual memory image when the verbal code is insufficient for correct identification. Both this and the verbal basis of recognition show up in color accuracy scores.

Our study asked a different question—were the distance relations which colors come to have to each other in memory influenced by the verbal code? Only the nature of memory errors provide data relevant to this question, that is, only cases in which verbal and/or visual memory codes were insufficient for completely correct identification. We predicted interaction between the verbal and visual storage. Such interaction could have resulted from various mechanisms: Since the visual image appears more "labile" than the verbal item, the verbal code might have become a fixed reference point for the image; that is, as the memory image faded or changed over time, it might have changed along lines "laid down" by the concurrent verbal code. Or the verbal code might have set limits for imagery changes; for example, an image tagged "mola" by a Dani S might have shifted in any direction so long as it remained within the color space recognized as "mola" by that S. Or the verbal code might have caused a "bias" at the point of response only. By any of these mechanisms, colors which were closer to each other in the name structure should have been more often confused with each other in recognition.

We did not find such an effect. Instead, the relation of colors to each other in memory was very similar to the way the color space appears to be structured in perception; that is, the circular structure of the hues of each brightness in our scaling solution was very similar to the structure obtained by Helm (1964) from similarity ratings of triads of ten visually present hues (of equal brightness and saturation) and to the structure obtained by Shepard (1962) from similarity ratings (reported by Ekman, 1954) of dyads of 14 visually present hues. We can say that the memory images in our task were isomorphic to the visual images of physically present colors, isomorphic in the sense (Shepard & Chipman, 1970) that the memory images bore the same relationship to each other that the visual images bore to each other. We do not know what the mechanisms are that operate to preserve isomorphism. We do know that, in the present task, they were quite resistant to language-related distortion. Descriptively, we can say that "mental" visual images, at least of colors, like "perception itself" (Gibson, 1967), do not appear easily changed by language. That is, visual coding appears to be a process separate from verbal rehearsal, a process which does not readily interact with verbal memory control processes. Perhaps whenever visual coding of the stimulus is not explicitly prevented by the design of a task, retention of visual images will be resistant to language related distortions.

In conclusion, we wish to argue that, although there are linguistic variables which correlate with color memory accuracy under certain conditions, the nature of color memory images themselves and the way in which they structure the color space in memory appear little influenced by language.

REFERENCES

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

BROWN, R., & LENNEBERG, E. A study in language and cognition. *Journal of Abnormal and Social Psychology,* 1954, **49**, 454–462.

BURNHAM, R., & CLARK, J. A test of hue memory. *Journal of Applied Psychology,* 1955, **39**, 164–172.

CLIFF, N. Orthogonal rotation to congruence. *Psychometrika,* 1966, **31**, 33–42.

EKMAN, G. Dimensions of color vision. *Journal of Psychology,* 1954, **38**, 467–474.

FARNSWORTH, D. *The Farnsworth–Munsell 100 hue test for the examination of color discrimination: manual.* Baltimore: Munsell Color Company, 1949.

GARNER, W. R., & FELFOLDY, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology,* 1970, **1**, 225–241.

GIBSON, E. J. *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts, 1967.

HEIDER, E. R. Universals in color naming and memory. *Journal of Experimental Psychology,* 1972, in press.

HEIDER, K. G. *The Dugum Dani: a Papuan culture in the Highlands of West New Guinea.* Chicago: Aldine, 1970.

HELM, C. E. Multidimensional ratio scaling analysis of perceived color relations. *Journal of the Optical Society of America,* 1964, **54**, 256–262.

KOEN, F. Codability of complex stimuli: three modes of representation. *Journal of Personality and Social Psychology,* 1966, **3**, 435–441.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika,* 1964, **29**, 1–27. (a)

KRUSKAL, J. B. Nonmetric multidimensional scaling: a numerical method. *Psychometrika,* 1964, **29**, 115–130. (b)

LANTZ, D., & STEFFLRE, V. Language and cognition revisited. *Journal of Abnormal and Social Psychology,* 1964, **69**, 472–481.

LENNEBERG, E. Color naming, color recognition, color discrimination: a re-appraisal. *Perceptual and Motor Skills,* 1961, **12**, 375–382.

NEISSER, U. *Cognitive Psychology.* New York: Appleton-Century-Crofts, 1967.

NEWHALL, S. M., NICKERSON, D., & JUDD, D. B. Final report of the OSA subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America,* 1943, **33**, 385–418.

NICKERSON, D., TOMASZEWSKI, J. J., & BOYD, T. F. Colorimetric specifications of Munsell repaints. *Journal of the Optical Society of America,* 1953, **43**, 163–171.

POSNER, M. Components of skilled performance. *Science,* 1966, **152**, 1712–1718.

POSNER, M. Abstraction and the process of recognition. In G. H. Bower, & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 3. New York: Academic Press, 1969.

SHEPARD, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function. I and II. *Psychometrika,* 1962, **27**, 125–140 and 219–246.

SHEPARD, R. N. Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology,* 1964, **1**, 54–87.

SHEPARD, R. N., & CHIPMAN, S. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology,* 1970, **1**, 1–17.

SCHÖNEMANN, P. H. A generalized solution of the orthogonal Procrustes problem. *Psychometrika,* 1966, **31**, 1–10.

SCHÖNEMANN, P. H., & CARROLL, R. M. Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrika,* 1970, **35**, 245–255.

STEFFLRE, V., CASTILLO VALES, V., & MORLEY, L. Language and cognition in Yucatan: a cross cultural replication. *Journal of Personality and Social Psychology,* 1966, 4, 112–115.

WHORF, B. L. *Language thought and reality.* Cambridge, Mass.: MIT Press, 1956.

(Accepted September 30, 1971)