A STUDY IN LANGUAGE AND COGNITION[1]

ROGER W. BROWN
Harvard University

AND ERIC H. LENNEBERG
Center for International Studies, Massachusetts Institute of Technology

It is popularly believed that reality is present in much the same form to all men of sound mind. There are objects like a house or a cat and qualities like red or wet and events like eating or singing and relationships like near to or between. Languages are itemized inventories of this reality. They differ, of course, in the sounds they employ, but the inventory is always the same. The esthetic predilections of the Italian lead him to prefer euphonious vowels, while the German is addicted to harsh consonant groupings, but the things and events named are the same in both tongues. We are confirmed in this view by our linguistic education, which requires us to memorize lists of French or German or Latin words and their exact English equivalents.

There are, of course, poetic persons who claim to find in each language some special genius that peculiarly fits it for the expression of certain ideas. But the majority of us are at a loss to understand how this can be, since there is apparently a relationship of mutual translatability among the languages we learn. To be sure, we can see that one language might contain a few items more than another. If the Germans were to invent a new kind of automobile and we had not yet thought of such a machine, their dictionary would have one entry more than ours until we heard of the discovery and named it for ourselves. But these inequalities are in the lexical fringe. They do not disturb the great core of common inventory.

The Whorf Thesis

This linguistic ethnocentrism will be seriously disturbed by the study of languages that lie outside the Indo-European group. It has not prepared us for finding that there is a language in which noun and verb categories apparently do not exist, or that there is another in which the colors gray and brown are called by the same name. Such data from the study of American Indian tongues led Whorf (18) to reject the usual view of the relationship between language and thought. He suggested that each language embodies and perpetuates a particular world view. The speakers of a language are partners to an agreement to see and think of the world in a certain way—not the only possible way. The world can be structured in many ways, and the language we learn as children directs the formation of our particular structure. Language is not a cloak following the contours of thought. Languages are molds into which infant minds are poured. Whorf thus departs from the common sense view in (*a*) holding that the world is differently experienced and conceived in different linguistic communities and (*b*) suggesting that language is causally related to these psychological differences.[2]

Other authors have believed that the relationship between language and thought is somewhat as proposed by Whorf. Cassirer (2), the distinguished philosopher, maintained that language is the direct manifestation of knowledge; he explicitly denied a form-content relationship between words or language structure and isolates of knowledge. In this he was in agreement with such other German writers as Wundt (19) and Bühler (1). Orwell (15) in his novel *Nineteen Eighty-Four* describes a totalitarian England of the future. The really efficient dictatorship of that day invents a language—Newspeak—in which it is impossible not only to express, but even to think, a rebellious thought. An equally great faith in the causal efficacy of language lies behind the General Semantics movement. Korzybski (8), for instance, holds that clear thinking and so-

[1] This project was carried out under the joint auspices of the Communications Program of the Center for International Studies, Massachusetts Institute of Technology, and the Laboratory of Social Relations, Harvard University.

[2] While this seems a fair statement of Whorf's usual views, he occasionally took a somewhat more conservative position.

cial progress are to be attained through the reform of language.

Cognitive Differences between Linguistic Communities

The first tenet of the Whorf thesis is that the world is differently experienced and conceived in different linguistic communities. The evidence presented in support of this claim is entirely linguistic. It will be helpful to distinguish between the conclusions based on lexical features of two languages and those based on structural features.

Lexical features. In the Eskimo lexicon there are three words to distinguish three varieties of snow. There are no single-word equivalents for these in English. The word "snow" would be used to describe all three. What psychological conclusions can be drawn from these data? Does the Eskimo see differences and similarities that we are unable to see?

Psychologists ordinarily infer perceptual discrimination when a subject is consistently able to respond differently to distinctive stimulus situations. The subject may be rat, dog, or man. The response may be running, salivation, or—speech. Words are used meaningfully when they are selectively employed with reference to some kind of environment—whether physical, social, or linguistic. The linguist in the field may discover the referent of a term by noting the pattern of its usage. The Eskimo's three "snows" are sufficient evidence from which to infer that he discriminates three varieties of snow. These selective verbal responses satisfy the conditions for inferring perceptual discrimination.

What can be said of the English speaker's ability to distinguish the same three kinds of snow? When different stimuli do not elicit differential responses, the stimuli may or may not be discriminated. A subject may be perfectly able to distinguish two situations and still not care to do anything about it. Consequently the fact that English speakers do not have different names for several kinds of snow cannot be taken to mean that they are *unable* to see the differences. It would seem, then, that all such comparisons are psychologically inconclusive. The Eskimo and American may or may not see the world differently.

There is, however, other evidence to indicate that the speaker of English can classify snows as the Eskimo does. If we listen to the talk of small boys, it is clear that they perceive at least two kinds of snow—good-packing snow and bad-packing snow. This is a distinction of the greatest importance to anyone interested in making snowballs. This discrimination is evidenced by differential response—not distinct lexical items but combinations of items—"good-packing snow" and "bad-packing snow." Whorf himself must have been able to see snow as the Eskimos did since his article describes and pictures the referents for the words. Since both Eskimo and American are able to make differential responses to snows, we must conclude that both are able to see differences. This seems to lead us to the conclusion that the Eskimo and American world views do not differ in this regard.

Although the three kinds of snow are namable in both Eskimo and English, each of them requires a phrase in ordinary English, whereas single words will do it for the Eskimo. Zipf (20) has shown that there exists a tendency in Peiping Chinese, Plautine Latin, and American and British English for the length of a word to be negatively correlated with its frequency of usage. This is true whether word length is measured in phonemes or syllables. It is not difficult to find examples of this relationship in English. New inventions are usually given long names of Greek or Latin derivation, but as the products become widely known and frequently used in conversation the linguistic community finds shorter tags for them. Thus the "automobile" becomes the "car" and "television" shrinks first to "video" and eventually to "TV." Three-dimensional movies are predictably described as "3-D."

Doob (3) has suggested that this principle bears on Whorf's thesis. Suppose we generalize the findings even beyond Zipf's formulation and propose that the length of a verbal expression provides an index of its frequency in speech and that this, in turn, is an index of the frequency with which the relevant perceptual judgments of difference and equivalence are made. If this is true, it would follow that the Eskimo distinguishes his three kinds of snow more often than Americans do. It would mean —to cite another example—that the Hopi is less often called upon to distinguish airplanes, aviators, and butterflies than is the American, since the Hopi has but a single name for all three of these. Such conclusions are, of course, supported by extralinguistic cultural analysis, which reveals the importance of snow in the Eskimo's life and the comparative indifference of the Hopi to airplanes and aviators.

We will go further and propose that increased frequency of a perceptual categorization will mean a generally greater "availability" of that category. In the experimental study of memory we are accustomed to think of the methods of recall, recognition, and relearning as increasingly sensitive indices of retention. In the experimental study of categorizing behavior there are two principal methods: (*a*) Goldstein's (6) technique of presenting a subject with an array of objects and asking him to group them, and (*b*) Hull's (7) discrimination learning technique. Hull's method seems to be the more sensitive of the two. We should guess that when the Eskimo steps from his igloo in the morning and is confronted by a snowy world, these snows will fall into named categories for him in a way that they will not for the American. If, however, the American were subjected to a discrimination learning experiment, or if the perceptual structure were otherwise made worth his while, he could see snow as does the Eskimo. We think, really, that more namable categories are nearer the top of the cognitive "deck."

Structural features. Members of structural categories have no phonetic common denominator. They are grouped together because they have the same structural relations with other forms in the language. In English, nouns constitute a structural category; its members can appear with definite and indefinite articles, can form plurals in certain ways, etc. In French all nouns of the feminine gender belong to one structural category since they all require the feminine articles and suffixes.

Whorf generally assumes that structural categories are also symbolic categories. When he finds structural differences in languages he concludes that there are parallel cognitive differences. There are in Hopi two structural categories showing some similarity to our verb and noun categories, with the difference that one of the Hopi classes includes only the names for such short-term events as lightning, flame, and spasm, while the other includes only such long-term events as man, house, and lifetime. Whorf concludes that the Hopi organizes his world on a dimension we usually overlook. When the structural class has such obvious semantic properties, Whorf's conclusions have a kind of plausibility.

However, very few structural classes have such clear and consistent meanings. In the languages we know best, those of the Indo-European family, there are many structural categories with no discernible meaning. In French, for instance, it is not clear that the gender of a form signifies anything to a speaker. Certainly it is difficult to find any common attributes in the references for French nouns of feminine gender. Not even the majority of them manifest feminine sexuality—even in the extended sense of Freud. The French speak of "*le* balcon" in spite of their saying, "Elle a du balcon." The linguist Charles Fries (5) has shown how difficult it is to describe a semantic for the English "parts of speech." If the noun can be defined as "the name of a person, place, or thing," this is only because "thing" is left unexplicated. It serves handily to designate whatever is nominalized and yet neither person nor place.

Even where the ethnolinguist can discover consistent structural meanings, it does not follow that these meanings are present to the native speakers of a language. Suppose that a subject in the laboratory were required to signal his recognition of each of ten different musical chords by raising that one of his ten fingers which has been designated for each chord. If all extraneous sensory information were excluded, his ability to pattern correctly the movements of his fingers would be evidence of his ability to identify the chords. The experimenter might introduce a potential structural meaning by ruling that the fingers of the right hand would always be raised for chords in the major mode and the fingers of the left hand for minor chords. The subject's responses might follow this pattern and yet he need never have detected the major and minor modes. Similarly, even if there were some semantic to French gender, one could speak the language without detecting it. "La fille" and "la femme" could be learned without noticing that both are in the feminine mode. No safe inferences about cognition can be made on the basis of the simple existence of the structural classes described by Whorf. The structural evidence is extremely difficult to interpret, and it seems clear that psychological data are needed to supplement those of the linguist.

Language in Causal Relation to Cognition

The second major tenet of Whorf's thesis is that language causes a particular cognitive structure. In what way can this occur? There seem to be two possibilities. Suppose that the colors red and green are not "given" categories but must be learned. A father who has formed these categories may play a game with his child that will teach the categories. The green blocks are to be used for building a house and the red ones for a barn. The child cannot properly pattern the blocks without learning to make the visual distinctions involved. Notice that the barn and house are not essential here. A father could ask his child to tell him whether each block is red or green. In learning this game, too, the child necessarily would learn to perceive the colors. Because words have symbolic properties, because their usage is patterned with reference to the total environment, language can cause a cognitive structure. To the degree that children are motivated to speak a language as it is spoken in their community they are motivated to share the world view of that community. To be sure, linguistic training is not the only means of procuring cognitive socialization; the house-barn game demonstrates that. The word game has the tremendous advantage that it can be played constantly and concurrently with many other activities. The child and his adult tutor can chatter together whether they are walking or riding, playing or working. In this chatter more is taught than a simple motor skill involving the muscles of articulation. A total culture is internalized.

There is a second, more dubious, avenue for the influence of language on thought. If life is a river, speech is a babbling brook whose course parallels that of the river. The brook is smaller and simpler than the river. A child can learn the phonemic structure of his language fairly easily. He will also realize that as the phonemic patterns he hears spoken change there are important changes in the nonlinguistic world. There is, for instance, an important difference that goes with the shift of speech from "father" to "mother." When, on the other hand, combinations of phonemes are repeated, two situations are equivalent in some important way. Consider the "strike" and the "ball" in baseball. These are rather difficult categories. The differences between them are subtle and complex. A naive observer of a baseball game would have a difficult time learning these categories by simply observing the game. It makes a great difference that the umpire calls out "strike!" each time a member of that category occurs and "ball!" to identify an instance of the other category. The umpire's shout directs us to look here and now to discover something of importance. The word spotlights a moment of consciousness and puts it in connection with other events similarly spotlighted. The various "strikes" are equivalent in some way and distinct as a category from the events labelled "ball." The babbling brook can, then, be a guide to the structure of the more complex but also more interesting river.

All of our reasoning cannot be said to prove the validity of any set of psychological conclusions. It does, however, point the direction for such a proof and suggests empirical steps that will advance our knowledge of this problem. We have made a small beginning in this work.

Our findings bear on only one of the claims made by Whorf—that there are cognitive differences correlated with lexical differences. We have developed lexical differences into the variable of "codability" and attempted to spell out the relationship between this variable and a single cognitive performance—recognition.

The Experiment

Sensory psychologists have described the world of color with a solid using three psychological dimensions: hue, brightness, and saturation. The color solid is divisible into millions of just noticeable differences; *Science of Color* (14) estimates 7,500,000. The largest collection (4, 11) of English color names runs to less than 4,000 entries, and of these only about 8 occur very commonly (17). Evidently there is considerable categorization of colors. It seems likely to us that all human beings with normal vision will be able to make approximately the same set of discriminations. This ability appears to depend on the visual system, which is standard equipment for the species. Whatever individual differences do exist are probably not related to culture, linguistic or extralinguistic. It does not follow that people everywhere either see or think of the color world in the same way. Cultural differences probably operate on the level of categorization rather than controlled laboratory discrimination.

Our explorations in the Yale Cross-Cultural Index turned up many reports of differences on this level. Seroshevskii (16), for instance, has reported that in the Iakuti language there is a single word for both green and blue. This is the kind of language difference discussed in the first section of this paper. A region of experience is lexically differentiated in one culture but undifferentiated in another. Color categories differ from such categories as snows in that they have boundaries that can be plotted on known dimensions. Color categories, furthermore, are continuous with one another, sharing their boundaries. Consider for a moment the single dimension of hue taken at a high level of saturation and brightness. Native speakers of English could be shown various shades and asked to give the usual color name for each stimulus presented. For each common color name there would be some shades invariably called by that name. There would be other shades sometimes associated with one name, sometimes with another. When the responses are divided about equally between two or more names, we should have boundaries between categories. If a native speaker of Iakuti were asked to provide the usual color names for the various shades, we should anticipate a somewhat different pattern. English speakers would have trouble naming the hues in the boundary region between green and blue. Probably they would hesitate, disagree among themselves, and sometimes use phrases or such combination names as "greenish blue." For the Iakuti, on the other hand, this region is right in the center of a category and would be named with great ease.

Of course, our example is greatly simplified over the actual case since we have dealt with the single dimension of hue whereas the color lexicon is actually patterned with respect to all of the three dimensions of visual experience. When these are considered, the range of applicability of a color term is a space within the color solid rather than a distance along a line. The simplification was for expository purposes and does not alter the logic of the argument.

This example of a cultural difference serves to introduce the variable *codability*. Certain colors are differentially codable in the Iakuti and English languages. So long as the data collected are of the usual linguistic variety, this difference of codability will be manifest in only one way—environmental distinctions expressed lexically in one language are expressed with word combinations in another language. Our reasoning led us to expect differential availability of reference categories in such a case. We undertook experimental work to discover additional behavioral indices of codability, and hoped to find one more sensitive than that which can be teased out of linguistic data. If we found such an index, we would go on to explore the behavioral consequences of differential availability of cognitive categories.

There are differences of codability within English itself. Some shades fall safely within the province of a given name while others lie within boundary regions. Here it is a matter of comparing the English codability of one region of visual experience with another region, whereas the ethnolinguist has usually compared the codability of one region of experience in several languages. If we explore the codability variable in English, it seems likely that our discoveries will apply to the cultural differences with which the inquiry began. If a general law can be found relating codability to availability, individual cultures may conform to the law though they differ among themselves in the values these variables assume in particular regions of experience.

Measurement of Codability

The entire series of Munsell colors for the highest level of saturation ("chroma" as Munsell calls it) was mounted on cards in systematic fashion. Five judges were asked to pick out the best red, orange, yellow, green, blue, purple, pink, and brown from these 240 colors. These names are the most frequently appearing color terms in English (17). For each name the color chip most often selected was added to our test list. Agreement among judges was high, and it is quite clear, therefore, that there is in this series one particular color chip with the best claim to each color name. The number of colors was then raised to 24 by adding chips that would, in combination with the first 8, provide as even a coverage of the color space as practicable. These colors are specified in Table 1. One set of the 24 chips was mounted on white 3 × 5 cards, one chip to a card. Another set was arranged randomly on a single large card.

To expose the single small cards a drop shutter was mounted in a 3 × 2-foot gray (Munsell neutral value 6, reflectance 30 per cent) board. The board was about three feet from the subject's (*S*'s) eyes and was

TABLE 1

The Munsell Notation and Scores for Discriminability, Codability, and Recognition for the 24 Test Colors

Munsell Notation*	Discriminability		Codability		Recognition (Group C Table 3)	
	Score	Rank	Score	Rank	Score	Rank
2.5R 7/8	38	2	18	9.5	.875	8
2.5R 5/10	27.5	6	7	18.5	.694	11
5R 4/14	23	10.5	19	7.5	1.020	5
7.5R 8/4	18	15	7	18.5	.236	18
2.5YR 6/14	38	2	29	1.5	1.499	2
5YR 3/4	24	9	26	3	.972	7
7.5YR 5/8	26	7.5	8	16	.736	9
2.5Y 7/10	12	19	3	24	.486	13
5Y 8/12	37	4	25	4	2.450	1
7.5Y 6/8	13	17	4	23	.250	17
3GY 7.5/11.2	23	10.5	14	12	1.222	4
7.5GY 3/4	9.5	23	14	12	0.000	23.5
2.5G 5/8	18.5	14	23	6	.986	6
7.5G 8/4	17.5	16	19	7.5	.167	19
5BG 3/6	4.5	24	12	15	.111	22
10BG 6/6	21	12	7	18.5	.458	14
8.5B 3/6.8	38	2	13	14	0.000	23.5
2.5PB 7/6	19	13	18	9.5	.436	16
5PB 4/10	10.5	21	29	1.5	.695	10
10PB 5/10	12	19	7	18.5	.125	20.5
5P 8/4	12	19	14	12	.547	12
10P 3/10	10	22	24	5	.444	15
5RP 6/10	26	7.5	6	21.5	.125	20.5
8RP 3.4/12.1	31	5	6	21.5	1.464	3

* For conversion to C.I.E. Tristimulus values and Source C, C.I.E. chromaticity coordinates see Nickerson, Tomaszewski, and Boyd (13).

illuminated from above and behind by a General Electric standard daylight fluorescent lamp.

The *S*s were 24 Harvard and Radcliffe students who spoke English as a native language and had no particular training in distinguishing colors. They were screened for color blindness with the standard Pseudo-Isochromatic Plates.

The *S*s were first shown the 24-color random chart for about five minutes. After the chart was removed, they were told that each of the colors on the chart would appear individually in the tachistoscope and that *S*'s task was to give the name of each as it appeared. "Name" was defined as the word or words one would ordinarily use to describe the color to a friend. The *S*s were urged to be both quick and accurate.

The 24 colors were presented in a predetermined random order for each *S*. No order was repeated. Each color was exposed until *S* had named it. In our trial procedure we used a voice key and chronoscope to measure the reaction time. The scope was activated by the opening shutter of the tachistoscope and stopped by *S*'s first vocalization. This method proved to be unsuitable since *S*s would frequently burst out with something other than a color name, which, of course, stopped the undiscriminating chronoscope. Consequently, we abandoned this technique and used the stop watch. The watch was started as the experimenter (*E*) dropped the shutter and stopped at first mention of a color name.

The variable of codability was measured in five ways. (*a*) The average length of naming response to each color was obtained by counting syllables. (*b*) The average length was also obtained by counting words. (*c*) The average reaction time for each color was obtained by ranking all of the reaction times of an individual *S* and taking the mean rank across *S*s for every color. (*d*) The degree to which *S*s agreed with one another in naming a color was assessed as follows: We counted the total number of different responses to a color (DR) and also the number of *S*s who agreed on whatever response was most often given to a particular color (CR). The first value was subtracted from the second and a constant of 20 added to keep the results positive (CR − DR + 20). Color 18, for example, was given the following eight different names: gray-blue, blue, light gray-blue, light blue, very pale blue, light blue-gray, pale blue, and powder blue. Of these, the single-word response "blue" occurred most often—six times. Color 18, then, scored 6 − 8 + 20, or 18. (*e*) The degree to which *S*s agreed with themselves from one time to another in naming a color was calculated as follows: Five *S*s were recalled after a period of one month and subjected to a repetition of the naming procedure. When an *S* gave identical responses to a color on the two occasions, we counted one agreement. We determined the number of agreements for each *S* and considered that to be unity. Each individual agreement was then given the appropriate fractional value. Suppose an *S* had eight agreements. If he agreed in his name for Color 11, he would add ⅛ to the score for that color. The agreement score is, then, the sum of the individual performances weighted for each individual's over-all tendency to agreement.

In Table 2 the intercorrelations of scores on these five measures appear. All correlations are in the predicted direction and most of them are significant, with .355 the smallest. With a single iteration this matrix yielded a general factor which we call codability. No correlations over .113 remain after the extraction of this single factor. Our fourth index, the degree of agreement between *S*s, has by far the largest factor loading. It was selected as the measure of codability for the second phase of the

TABLE 2

Correlation Matrix for Five Indices of Codability

Measure	1	2	3	4	5
Number of syllables					
Number of words	.425*				
Reaction time	.387	.368			
Interpersonal agreement	.630*	.486*	.864*		
Intrapersonal agreement	.355	.537*	.649*	.773*	
k from second factoring	.589	.587	.787	.976	.795
Communality from first factoring	.403	.378	.671	.873	.653

* $p \leqq .05$.

experiment. The obtained codability values for the 24 colors are listed in Table 1.

Codability and Recognition

Once the codability variable suggested by Whorf's ethnolinguistic observations had been operationalized, it remained to relate this variable to some nonlinguistic behavior which might be considered an index of availability. We selected the recognition of colors.

From the 240 Munsell chips taken at highest saturation we selected out alternate chips, taking care to include the 24 colors for which codability data had been collected. The resultant collection of 120 colors was systematically mounted on a white card. Hue varied along the vertical dimension of the card and brightness on the horizontal dimension. Since there were 20 steps of hue and only 6 of brightness, we divided the total colors in half and mounted one half above the other so as to make a more manipulable display.

New *S*s were screened, as before, for color blindness and language background. The basic procedure was to expose simultaneously 4 of the 24 colors, remove them, and ask *S*s to point to the colors just seen, on the large chart of 120. Neither *E* nor *S* mentioned any color name during the session. The recognition score for a color was computed as follows: We determined the number of correct identifications made by each *S* and considered this number to be unity. Each individual correct identification was given the appropriate fractional value. Suppose for instance, that an *S* who correctly identified a total of six colors recognized Color 24. This recognition would have counted as 1/6 on the total recognition score for that color. Another *S* for whom Color 24 was one of eight correctly identified colors would have contributed 1/8 to the score for Color 24. In other words, the recognition score for a color is the sum of the individual performances weighted for each *S*'s over-all ability to recognize colors. The scores for the 24 colors appear in Table 1.

In trial runs, *S*s were asked how they managed to retain the four colors in memory after they were removed from sight. Most *S*s reported that they named the colors when they were exposed and "stored" the names. It seemed likely, therefore, that those colors that are quite unequivocally coded would be most likely to be recognized. When a color elicits a considerable range of names, the chances of recovering the color from the name would be reduced. This expectation was fulfilled by a rank-order correlation of .415 between codability and recognition scores.

There is, however, another variable that influenced recognition. The 120 colors used are not perceptually perfectly equidistant. The manufacture of equidistant color chips is technically difficult and expensive and, indeed, above a certain level of saturation, impossible. Since we were unable to control experimentally the variable "discriminability," we must ask whether or not our findings were due to a positive correlation between codability and discriminability. Could it be that our codable colors were so distant, perceptually, from their nearest neighbors that their superior recognizability was actually due to these better discrimination conditions? To obtain an answer to this question we determined the true perceptual distance between each of the colors used from the Newhall, Nickerson, and Judd (12) charts. These charts convert every Munsell book notation into a renotation which is the specification of a true perceptual locus of each color within the Munsell coordinate system. The difference between two renotations expresses quantitatively the perceptual distance between the colors.

For each of the 24 test colors we computed a discriminability score which describes its distinctiveness from the colors surrounding it. The difference between two renotations yields three numbers, one for each dimension. To make these numbers perceptually commensurable (i.e., to reduce them to a common denominator), the Optical Society of America Subcommittee on the Spacing of the Munsell Colors suggests the values 3, 2, and 1 for hue, chroma, and value, respectively. Since every color has two neighbors on each of the three dimensions, a total of six numbers will express, in a rough way, the discriminability of that color. The sum of these yields the unadjusted discriminability score. Adjustments of this score are necessary (*a*) because if a color appears on the margin of our chart it has a lower chance of being recognized correctly and (*b*) because a color that has a very close neighbor on one side and distant neighbors on three others might come out with a good discriminability score although the close contiguity on one side would hinder correct recognition considerably. Consequently, colors appearing on the margin of our chart had the constant 3 subtracted from their unadjusted discriminability score, and colors with a close neighbor had the constant 6 subtracted.

Our scoring method is to a certain degree arbitrary, to be sure, but since the equation of perceptual distances on different visual dimensions is an unsolved problem, there seems to be no more objective method available. In addition, of course, all decisions were made without knowledge of recognition scores.

Since we were unable to control discriminability experimentally, we controlled it statistically. The partial correlation between codability and recognition, with discriminability constant, is .438. Furthermore, the correlation between codability and discriminability is .074, which is not significant. Evidently the relation between codability and recognition is not a consequence of variations in discriminability.

Since the reports of our early *S*s indicated that colors were stored in linguistic code, it seemed likely that color codability would increase in importance as the storage factor was maximized in the recognition situation. Discriminability, on the other hand, should remain at the same level of importance or possibly decline somewhat. If, for example, a single color were exposed, removed, and then identified with minimal delay, *S*s

TABLE 3

RECOGNITION PROCEDURES

GROUP	*N*	NUMBER OF COLORS ORIGINALLY EXPOSED	LENGTH OF INTERVAL	CONTENT OF INTERVAL
A	9	1	7 seconds	
B	9	4	7 seconds	
C	16	4	30 seconds	
D	9	4	3 minutes	Tasks

Note.—Exposure time for all groups was 3 seconds.

TABLE 4

Correlations Involving Scores on Codability (C), Discriminability (D), and Recognition (R) with Four Experimental Conditions for Recognition

Group	C with R	D with R	C with R, D constant
A	.248	.540*	.248
B	.411	.460*	.426*
C	.415	.503*	.438*
D	.487*	.505*	.523*

* $p \leqq .05$.

might retain some direct memory of the color, perhaps as a visual image. In this situation discriminability would be a determinant of recognition but codability would not be. However, when the number of colors is increased and the interval prolonged and filled with activity, the importance of linguistic coding should increase. Table 3 describes the experimental variations we used. Groups A, B, C, and D are arranged in what we believed to be an order of increasingly difficult storage of colors. Group C is our major group, for which results have already been described. The tasks which filled the interval for Group D were simple but absorbing—the kind of thing often used in experiments on the Zeigarnik phenomenon.

It can be seen from the data in Table 4 that the correlation between recognition and codability scores does increase as the importance of storage in the recognition task increases.[3] The particular order obtained would occur by chance only once in 24 times.

Table 4 also shows that discriminability is most closely related to recognition in Group A, for which the possibility of some direct memory of the color is maximized. The importance of discriminability declines slightly but not significantly as the recognition is made more difficult. Our expectations with regard to both codability and discriminability are generally confirmed.

In the first section of this paper we concluded our discussion of lexical differences between languages with the prediction that a given set of cognitive categories will be more available to the speakers of a language that lexically codes these categories than to the speakers of a language in which the categories are not represented in the lexicon. Lexical differences have been expanded into the variable of codability, and category availability has been operationalized as a recognition score. We found that differences in the English codability of colors are related to differences in the recognition of these colors. We expected these results to apply to the cross-cultural case, and some confirmation of this expectation is available in the results of a study by Lenneberg and Roberts (10). This study of Zuni Indians used a field adaptation of our methods and apparatus. The Zuni color lexicon codes the colors we call orange and yellow with a single term. Monolingual Zuni *S*s in their recognition task frequently confused the orange and yellow colors in our stimulus set. Our English-speaking *S*s never made this error. It is a distinction which is highly codable in English and highly uncodable in Zuni. Interestingly, bilingual Zunis who knew English fell between the monolingual Zuni and the native speaker of English in the frequency with which they made these errors.

The Whorf thesis claims more than a simple relationship between language and cognition; language is held to be causally related to cognitive structure. Our correlational evidence does not, of course, establish the direction of causality. If we may be permitted a guess it is that in the history of a culture the peculiar features of the language and thought of a people probably develop together.

In the history of an individual born into a linguistic community the story is quite different. The patterned responses are all about him. They exist before he has the cognitive structure that will enable him to pattern his behavior in the approved fashion. Simple exposure to speech will not shape anyone's mind. To the degree that the unacculturated individual is motivated to learn the language of a community, to the degree that he uses its structure as a guide to reality, language can assume a formative role.

Summary

The Whorf thesis on the relationship between language and thought is found to involve the following two propositions: (*a*) Different linguistic communities perceive and conceive reality in different ways. (*b*) The language spoken in a community helps to shape the cognitive structure of the individuals speaking that language. The evidence for the first proposition derives from a comparison of the lexical and structural characteristics of various languages. The linguistic comparisons alone do not establish the proposition. They need to be complemented with psychological data. The

[3] Kurtz and Hovland (9) have shown that verbalization during observation of stimulus objects facilitates recognition under certain circumstances.

second proposition is not directly supported by any data. However, it is clear that language can be described as a molder of thought since speech is a patterned response that is learned only when the governing cognitive patterns have been grasped. It is also possible that the lexical structure of the speech he hears guides the infant in categorizing his environment. These matters require empirical exploration.

An experiment is described which investigates a part of proposition *a*—the idea that lexical differences are indicative of cognitive differences. Whorf reports many cases in which a given range of experience is lexically differentiated in one language whereas the same discriminations can only be described with phrases in another language. Rather than compare members of different linguistic communities, we chose to work with native speakers of English and to compare their linguistic coding of two regions of experience. Within the realm of color vision there are colors that can be named with a single word and others that require a phrase. This kind of linguistic difference in the length of name (measured by words or syllables) was found to be correlated with the latency of the naming response and the reliability of the response from person to person within the linguistic community and from time to time in one person. A factor analysis of these measures yielded a single general factor—codability. The measure carrying the largest factor loading was the reliability of naming response between individuals who speak the same language. This variable—the codability of a color—proved to be related to *S*s' ability to recognize colors. Codability accounted for more variance in the recognition task as the task was delayed and complicated to increase the importance of the storage factor. Data obtained from the Zuni Indians show a similar relationship between codability and recognition. It is suggested that there may be general laws relating codability to cognitive processes. All cultures could conform to these laws although they differ among themselves in the values the variables assume in particular regions of experience.

REFERENCES

1. BÜHLER, K. *Sprachtheorie.* Jena: G. Fischer, 1934.
2. CASSIRER, E. *The philosophy of symbolic forms.* Vol. 1. *Language.* New Haven: Yale Univer. Press, 1953.
3. DOOB, L. W. *Social psychology.* New York: Holt, 1952.
4. EVANS, R. M. *An introduction to color.* New York: Wiley, 1948.
5. FRIES, C. C. *The structure of English.* New York: Harcourt, Brace, 1952.
6. GOLDSTEIN, K. *Language and language disturbances.* New York: Grune & Stratton, 1948.
7. HULL, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**, No. 1 (Whole No. 123).
8. KORZYBSKI, A. The role of language in the perceptual processes. In R. R. Blake & G. V. Ramsey (Eds.), *Perception; an approach to personality.* New York: Ronald, 1951. Pp. 170–205.
9. KURTZ, K. H., & HOVLAND, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, **45**, 157–164.
10. LENNEBERG, E. H., & ROBERTS, J. M. The denotata of color terms. Paper read at Linguistic Society of America, Bloomington, Indiana, August, 1953.
11. MAERZ, A., & PAUL, M. R. *A dictionary of color.* New York: McGraw-Hill, 1930.
12. NEWHALL, S. M., NICKERSON, D., & JUDD, D. B. Final report of the OSA subcommittee on the spacing of the Munsell colors. *J. opt. Soc. Amer.*, 1943, **33**, 385–418.
13. NICKERSON, D., TOMASZEWSKI, J. J., & BOYD, T. F. Colorimetric specifications of Munsell repaints. *J. opt. Soc. Amer.*, 1953, **43**, 163–171.
14. OPTICAL SOCIETY OF AMERICA, COMMITTEE ON COLORIMETRY. *The science of color.* New York: Crowell, 1953.
15. ORWELL, G. *Nineteen eighty-four.* New York: Harcourt, Brace, 1949.
16. SEROSHEVSKII, V. R. *Iakuti.* St. Petersburg: Royal Geographical Society, 1896.
17. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., 1944.
18. WHORF, B. L. *Four articles on metalinguistics.* Washington: Foreign Service Institute, 1950.
19. WUNDT, W. *Völkerpsychologie.* Vol. 1. *Die Sprache.* Leipzig: Engelmann, 1900.
20. ZIPF, G. K. *The psycho-biology of language.* Boston: Houghton Mifflin, 1935.

Received April 12, 1954. Prior publication.