DOCUMENT RESUME

ED 021 228 48 AL 001 335

By- Dale, Philip Scott

CHILDREN'S COLOR CATEGORIES AND THE PROBLEM OF LANGUAGE AND COGNITION.

Michigan Univ., Ann Arbor. Center for Research on Language and Language Behavior.
Spons Agency- Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No- BR-6-1784
Pub Date 1 Feb 68
Contract- OEC-3-6-061784-0508
Note- 67p.; Report included in Studies in Language and Language Behavior, Progress Report No. VI.
EDRS Price MF-$0.50 HC-$2.76
Descriptors- *CHILD DEVELOPMENT, COGNITIVE PROCESSES, *MEDIATION THEORY, *MODELS, PERCEPTION, *PSYCHOLINGUISTICS, RECOGNITION, THOUGHT PROCESSES, VERBAL TESTS
Identifiers- Categorial Perception, *Color Categories, Whorf Hypothesis

Twenty-four four-year-old children were given a perceptual task (color matching), a memory task (color recognition), and a verbal task (color naming). The relation between the children's performance on the nonverbal and verbal tasks was demonstrated by the fact that nearly all the matching and recognition responses, even when incorrect, were from the same naming class as the stimulus color. Contrary to expectations, this tendency was stronger on the matching task than on the recognition task. Only those children who spontaneously named the colors as they were presented showed this effect on the recognition task. As a result of this experiment, a model is proposed in which perception of colors leads to two distinct encodings: one representational and one symbolic. The two can have different internal histories, but if the symbolic encoding is present at the time of response selection, it will influence the output. However, the symbolic encoding is very short-lived for these young children, and it will be forgotten unless rehearsed overtly. This model is considered in terms of mediation; and it is argued that the non-namers in the recognition task do not suffer from production deficiency, but rather from mediational deficiency, resulting from a lack of memory for the mediator. These conclusions differ from the weak form of the Whorf hypothesis in that two distinct, concurrent encodings are assumed. (DO)

BR-6-1784
PA-48

CHILDREN'S COLOR CATEGORIES AND THE PROBLEM OF LANGUAGE AND COGNITION[1]

Philip Scott Dale

Center for Research on Language and Language Behavior
The University of Michigan

Young children, whose language is in a state of development and change, are particularly appropriate for the investigation of the relation between language and cognition, and the well-known Whorf hypothesis. Although previous measures, such as codability and communication accuracy, are not suitable, a simpler method, based on the classes of colors to which the subject gives the same name, is effective.

In this experiment, children approximately 4 years old were given a perceptual task (color matching), a memory task (color recognition), and a verbal task (color naming). The relation between the children's performance on the nonverbal and verbal tasks was demonstrated by the fact that nearly all the matching and recognition responses, even when incorrect, were from the same naming class as the stimulus color. Contrary to expectations, this tendency was stronger on the matching task than on the recognition task. Only those children who spontaneously named the colors as they were presented showed this effect on the recognition task.

It is argued that this is neither categorial perception, nor a shift of the entire distribution of responses towards the center of the category, but a simple inhibition of extra-categorial responses. A model is proposed in which perception of colors leads to 2 distinct encodings: one representational and one symbolic. The 2 can have different internal histories, but if the symbolic encoding is present at the time of response selection, it will influence the output. However, the symbolic encoding is very short-lived for these young children, and it will be forgotten, unless it is rehearsed overtly.

This model is considered in terms of mediation, and it is argued that the non-namers in the recognition task do not suffer from production deficiency, but rather from mediational deficiency, resulting from lack of memory for the mediator.

These conclusions differ from the weak form of the Whorf hypothesis in that 2 distinct, but concurrent encodings are postulated. In contrast, the weak form claims that the production of language makes certain aspects of the situation more salient, and leads to a biased, but still unitary, representational storage.

All the children appear to be producing the mediator, which is some form of category identification. Some, however, cannot remember it. The only mechanism they have to help them remember is rehearsal, and for children of this age, rehearsal must be overt. As they grow older, they acquire the ability to rehearse internally. Thus what becomes internalized is not speech, as Vygotsky claimed, but rather the ability to pay attention to inner speech, and in particular, to remember it by rehearsing.

[1]This dissertation was supported in part by the Language Development Section, U.S. Office of Education, Contract OEC-3-6-061784-0508, and it has been made a part of Studies in language and language behavior, Progress Report VI, February 1, 1968.

CHAPTER I - The Problem

Introduction

> 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.
>
> --Brown and Lenneberg (1954), p. 454

Nevertheless, a consideration of non-Indo-European languages, especially American Indian languages, led Whorf to a radically different view:

> We are thus introduced to a new principle of relativity, which holds that all observers are not led by the same physical evidence to the same pictures of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated.
>
> --Whorf (1956), p. 214

Whorf claimed that the world is experienced differently in different linguistic communities, and that this is caused by the language differences. He felt that even more important than the vocabulary incongruences known to all those familiar with two languages, e.g. English *piece* versus French *piece* and *morceau*, were differences in the grammatical structures of languages, which lead the language user to dissect the world along the lines specified in the grammar.

The so-called Whorf hypothesis (or Whorf-Sapir hypothesis, for the linguist Sapir shared in the development of the ideas) has been shown to be both ambiguous and partially untestable. Whorf's own use of translation as a method of investigation has been criticized effectively by Lenneberg (1953). Most of the anthropological attempts to verify the hypothesis have proceeded by considering some portion of the linguistic system of a culture and some corresponding portion of the non-linguistic "system," and then attempting to demonstrate some similarity between the two. For example, Hoijer (1964) attempted to relate more or less independent observations of Navaho verb categories and the Navaho view of movement and action, while Mathiot (1964) sought and found semantic correlates of Papago syntactic form-classes which agreed with folk taxonomies of plants, animals, and birds. These are essentially attempts to determine a correlation with just one data point. And as such, they are unconvincing.

A few cross-cultural experiments have been performed to overcome this limitation. One of these was done with Navaho-dominant and English-dominant Navaho children by Carroll and Casagrande (1958). Navaho verbs of handling require an inflection appropriate to the shape of the object handled: presumably this leads to an increased perceptual saliency of the properties so encoded. Carroll and Casagrande presented triplets of objects to the children, and asked them to select the pair of objects that were most similar. One triplet consisted of a blue rope, a yellow rope, and a yellow ball; other triplets were constructed analogously, so that one pair of objects shared a color, while another pair shared a shape. The experimenters found that Navaho-dominant Navahos paired objects taking the same verb suffix twice as often as English-dominant Navahos; apparently the Navaho-dominant children, but not the English-dominant ones, were guided by the verb system. However, when a group of New England children, all monolingual in English, were given the task, they made choices more often in the direction predicted for the Navahos than the Navaho-dominant children did. This makes the interpretation of the results of the experiment with Navaho children highly suspect.

This outcome is also a good example of the problems of cross-cultural experimentation in this area. Subjects in two different linguistic communities are separated by more than their languages: much of their culture, attitudes, experience, and interests will be different. This is also true, though to a lesser extent, of different linguistic groups within a single society, as in the case of the two groups of Navahos. But it is precisely the relation between language and these other factors that is the point of interest, so in general no inferences can be made. One solution to this problem is to restrict experimentation to the "language of experience," which will be discussed in the third section of this chapter.

Refinement of the Hypothesis

Several attempts have been made to systematize the Whorf hypothesis, and separate the various claims that are often confused in the writings of Whorf and others (Fishman, 1960; McNeill, 1965). McNeill's division of the hypothesis into three forms will be considered here.

McNeill illustrates the difference between the strong and weak versions of the hypothesis with the example of Navaho verb inflections mentioned earlier. One such inflection is the suffix *-ti;h*, denoting that something long and slender is the subject of the verb. There are twelve such verbal suffixes, and the presence of one of them is obligatory for all verbs of handling. If the object of the verb were a word such as the Navaho word for *stick*, then *-ti;h* would have to be included in the composition of the verb. The strong version claims that a general category of shape is established in the Navaho mind, a cognitive category derived from the linguistic category of those words requiring *-ti;h*.

All cognitive categories are held to be derived from linguistic categories in this fashion; cognition is patterned after language. For once the cognitive categories have been developed from language, they continue to exist, and form the basic material of cognition. The essential implication of this strong version is that the influence of language will hold in areas of cognition that are free of actual linguistic performance.

The weak version of the hypothesis, on the other hand, is not concerned with cognitive categories. It claims only that language may be of great use in non-verbal tasks, because it makes certain properties, namely, those encoded by the language, available for other cognitive purposes. But this only takes place when language is actually produced (which, of course, may be internally). Both the strong and weak form of the hypothesis predict that Navaho-dominant Navahos should be more likely to pair objects of the same shape in the Carroll and Casagrande experiment, but for different reasons. The strong form claims that cognitive categories of objects, based on their shape, already exist in the Navaho mind, derived from the Navaho language, and these will lead the Navaho to pair objects on the basis of shape. Under the weak form of the hypothesis, the same prediction is based on the assumption that a Navaho speaker actually uses the verbal suffixes in speaking to himself as he performs the task. Thus production is indispensable to the weak form, but not to the strong form. The Carroll and Casagrande study, then, did not differentiate the two forms of the hypothesis; but as has been shown, the results are inconclusive.

The weakest form of the hypothesis holds that the influence of language is limited to memory. The categories derived from language are, in general, not used for other cognitive purposes, except when it is necessary to remember something. The evidence for the weakest form (and the motivation for formulating it) comes from a series of experiments in color recognition memory, which will be discussed in the next section. McNeill summarizes the three versions of the hypothesis by contrasting the point of linguistic influence on cognition: "Whereas the strong version deals with thought, and the weak version deals with perception, the weakest version deals with memory." (McNeill, 1965, p. A40.)

In general, the weaker the form of the hypothesis, the more evidence there is for it.

The Color Experiments, Codability, and Communication Accuracy

Lenneberg and Roberts (1956) suggested that, for empirical studies of the relationship between words and their referents, attention should be restricted to words that have simple referents, in that they can be exhaustively described by their physical properties. This is only possible for that small class of words that describe the sensation of physical properties themselves. These are referred to as the *language of experience*. In general, the referents of these words are continuous in nature, and are classified into closed classes by their names. Examples are words for temperature, taste, and vision. Usually they are ordered along a small number of dimensions; one in the case of temperature, three in the case of color.

The first of a series of experiments studying the influence of language on color memory and recognition was performed by Brown and Lenneberg (1954). From a consideration of various languages, they hypothesized that colors are differentially *codable* in different languages. For example, although English speakers have difficulty naming colors in the region between blue and green, Iakuti speakers should not have such difficulty, because this region is in the center of the category corresponding to a color term in their language. Similarly, colors differ in their codability within a single language. Colors that lie near the boundary of a naming category are less codable than those lying well inside the category. In the first part of their experiment, Brown and Lenneberg measured codability with a number of indices: length of name given to color, naming latency, inter-subject agreement, and others. Their measures correlated highly, supporting the notion of a general codability factor.

It was hypothesized that codability would be correlated with recognition accuracy; that the easier it was to name a color, the easier it would be to remember it. In the second part of their experiment, Brown and Lenneberg measured recognition accuracy. One or more colors were presented to the subject, then they were removed, and after an interval of time the subject was asked to select the color seen earlier from a large array of colors. The pattern of the results was clear: when the task was easy (one color to be remembered for seven seconds), codability was not highly correlated with recognition accuracy; but as the task became more difficult the two began to correlate, and the relation was strongest when the task was most difficult (four colors to be remembered for three minutes).

Consideration of this experiment led McNeill to formulate the weakest form of the hypothesis, discussed above: the influence of language is essentially limited to memory. This is supported by the evidence of subjects who often reported that they named the colors when they were presented, and then stored the names.

Brown and Lenneberg also report a study of Lenneberg and Roberts. They administered the same experiment to Zuni Indians, and obtained similar results, despite the considerable difference in color terminology between the two languages. For example, the colors called orange and yellow in English are coded with a single term in Zuni. Monolingual Zunis frequently confused orange and yellow colors in the recognition task, though English speaking subjects never made this kind of error.

Burnham and Clark (1955) conducted a similar experiment, but using the Farnsworth-Munsell colors (Farnsworth, 1943) instead of the selection of high saturation colors used by Brown and Lenneberg. The Farnsworth-Munsell colors were used for two reasons: first, the series of colors had been constructed in such a way as to make each successive pair of colors perceptually equidistant (the lack of such equidistance had raised problems of analysis in Brown and Lenneberg's experiment): and second, the series form a circular stimulus array, eliminating endpoints. The recognition accuracy of the colors in this experiment was negatively correlated with codability, in contrast to the positive correlation obtained by Brown and Lenneberg. Both experiments have been replicated, with similar outcomes.

This apparent contradiction was resolved by Lenneberg (1961). The two stimulus arrays used by Brown and Lenneberg, and Burnham and Clark, differed in the fashion in which they were mapped by names. In the first array, one color corresponded well to each major color name in English, while the remainder were closer to boundaries. Thus a highly codable color, one well within the category, could be recognized easily, since it was the only color corresponding to its name. But "in-between-ness" was shared by many colors, and such colors could not be recognized easily. In this case, codability correlated positively with recognition accuracy. But in the second array, there were often several colors well inside any category, so, e.g., "green-ness" was no help in memory. But those few colors which fell near a boundary acquired a distinctiveness, and could be recognized on that basis. Codable colors, then, could not be recognized as well as less codable ones, and codability correlated negatively with recognition.

This analysis showed clearly the need for a better linguistic measure to predict nonverbal behavior. Such a measure was proposed and tested by Lantz and Stefflre (1964). They suggested that memory be viewed as a process in which an individual communicates to himself through time using the brain as a channel. This process, they claimed, is similar to communication between separate individuals. If an item can be communicated successfully to another individual, it can be remembered well. Their measure, called communication accuracy, was obtained by asking subjects to make up a description of a color that would enable another person to pick it out of an array of colors. Then these descriptions were given to another set of subjects to use in selecting colors. In this way, they measured the accuracy with which each item was communicated by the descriptions composed by the subjects. Lantz and Stefflre then performed the recognition experiment with both the stimulus arrays discussed above. Communication accuracy correlated positively with recognition on both arrays, while codability correlated positively with recognition accuracy for the Brown and Lenneberg array, and negatively for the Farnsworth-Munsell series. This experiment has been repeated with deaf and hearing adults, and hearing children with similar results (Lantz and Lenneberg, 1966). Communication accuracy appears to be an excellent predictor of recognition accuracy.

As might be expected on the basis of the Brown and Lenneberg study, the correlation between communication accuracy and recognition accuracy in the Lantz and Stefflre study was weakest for the easiest task (one color for five seconds) and stronger for the more difficult tasks (four colors for five and thirty seconds). This experiment, then, is evidence for only the weakest form of the hypothesis.

As a technique for detecting the influence of verbal behavior on nonverbal tasks, communication accuracy has certain limitations. First, it cannot be obtained in some cases, e.g., deaf children, children younger than the six-year-olds used in Lantz and Lenneberg. Second, it assumes a similarity between individuals that may not always hold. It may be valid to consider a process in which an individual communicates with himself through time, but this communication system may vary greatly from individual to individual. This is particularly likely in the case of very young children, whose color terms are often highly idiosyncratic.

For these reasons, a new technique for detecting this influence will be proposed in the third chapter of this thesis, one which considers each child's behavior, verbal and nonverbal, individually. The principle of the technique is as follows: if a subject is naming a color as input to the communication channel, then incorrect color selections are most likely to be other colors which he would have given the same color name.

Categorial Perception

Another relevant line of research is concerned more directly with the relation of linguistic categories and perception. It has been demonstrated by researchers in speech perception that as stimuli vary along certain continua, they are "heard" clearly as one of two (or more) categories. For example, if appropriate consonantal sounds are synthesized which vary only in the time of onset of the first formant, each sound will be heard as a /d/ or a /t/. The continuum is divided into discrete sets. It has also been found that discrimination is superior near the boundaries between these discrete categories; in fact, the discriminability of two stimuli can be predicted by the probability that the two stimuli are categorized differently (see Lane, 1967, for details). This phenomenon of enhanced discrimination near category boundaries is called categorial perception.

Kopp (1967) undertook an investigation to determine if categorial perception occurred in color vision. Using monochromatic stimuli, he measured hue labelling and latency, and hue discrimination, with the ABX and sweep techniques, for English speakers and Tzotzil speakers (in Mexico). He found discrimination peaks and longer latencies at naming category boundaries as predicted for the English speaking subjects, but a much less clear-cut relationship for the Tzotzil speakers.

A process opposite to categorial perception does not seem to have been considered in the literature. In this process, which will be termed attractive perception, stimuli well within a category are not affected, but those lying near the boundary tend to be drawn into one category or the other, according to their position. This process leads to a prediction of <u>inferior</u> discrimination at the boundaries, in contrast to categorial perception, which predicts <u>superior</u> discrimination there.

The Present Study

One solution to the need for data from more than a single homogenous group of language users is to study children. As children develop, they speak a sequence of at least partially distinct languages. This introduction of development as an independent variable allows inferences that cannot be drawn on the basis of observations of adult speakers of one language; including, to a limited extent, causality. Because the children can be selected from a single community, many of the problems of cross-cultural research can be avoided. For these reasons, the experiment reported in the following chapters was performed with children approximately four years old. The basic plan of research was to observe children's behavior on verbal and nonverbal color tasks; then first to demonstrate a relation between the two, and next to determine how this relation varied with the task, color, subject, and other aspects that might be relevant.

The verbal behavior observed was of two kinds. In the first, the child named the colors as they were presented to him. In the second, the child was given a name and asked to find the corresponding colors. These two approaches differ in important ways; for example, the second method will leave some parts of the color space unmapped, i.e., some colors will not be produced as responses to any color term. But this will not happen with the first method; in general, subjects will give some descriptive term to each color. Adult speakers of a language , and many children, can do both these tasks, and their behavior on the two will be related. If a color is called green, it is likely that it will be selected in response to the name green. Further, adults are aware that certain color terms describe some colors more accurately than other colors, and that the naming categories grade into each other (Lenneberg, 1967, p. 339; also, see Figure 3). These considerations led to the hypothesis that one of the two naming processes might be reducible to the other; this hypothesis was a secondary question of interest in the experiment.

Perhaps the fundamental act is that of naming a color. The most appropriate name would be produced first, but other terms might be produced. In this case, selecting colors in response to a name would be done by examining each color, naming it, and checking the names thus produced to determine if one matched the name given. The most important implication of this model is that naming categories produced by the second method should be larger than those produced by the first method. A color best described as green, but also describable as blue, would not be called blue, but would be produced in response to blue.

The opposite assumption is that the fundamental act is that of determining if a color is, e.g., green, and if so, if it is unambiguously so. In this case, naming a color requires the subject to check the color against each color word in his vocabulary until a match is found. Since subjects virtually always produce a name, it must be assumed that they do so even if the color is not unambiguously so described. An implication of this model is that naming categories produced by the second method should be smaller than those produced by the first. A color best described as orange, but also describable as yellow, would not be produced in response to the name orange, but would be named orange.

The nonverbal behavior was also of two kinds. An appropriate version of the recognition task was an example of a memory task. A more directly perceptual task was also included, in order to be able to compare it with the recognition task, and so test the strong and weak versus the weakest form of the Whorf hypothesis.

CHAPTER II - The Experiment

Introduction

The purpose of this experiment was to obtain independent observations of color behavior on three separate tasks: one perceptual, one memory, and one verbal. A secondary goal was the comparison of the results with earlier studies: this influenced the choice of stimulus materials.

Materials

Every sixth color chip (nos. 1, 7, 13, ..., 79) from the Farnsworth-Munsell series (Farnsworth, 1943) was purchased from the Munsell Color Company, Baltimore, Md. These chips, which have been used in most of the color experiments, are perceptually equidistant with respect to the sorting task by which the series was constructed. This equidistance allows an analysis of the magnitude of errors. However, the chips have the disadvantage of being of rather low saturation and brightness. There is, for example, no good instance of a "red" chip. Lantz and Lenneberg (1966) used every fourth chip for their experiment with six year olds; using every sixth chip makes the tasks easier for the younger children used in this study. For the purposes of this experiment, the chips were renumbered one through fourteen.

The color chips were covered with protective clear plastic discs and mounted in black plastic caps. They were arranged in a circular array on a white board.

Matching Task

Tests of perception with adults are often based on "same-or-different" judgements; that is, two colors are placed next to each other and one is adjusted until the subject indicates they appear identical to him. This technique was not used here, as the absolute limits of discrimination in children are not the point of interest. The literature reviewed by Rivoire and Kidd (1966) indicated that some discrimination ability exists in infants as young as 15 days, and pre-schoolers' ability is already quite similar to that of adults. In addition, the author has observed (in a few children) that if they are specifically requested to place a color chip next to "the chip with the very same color," and they are further motivated by the question, "are these two colors just the same?" when they offer a response, they will be able to do the task perfectly.

If the task is less structured, i.e., the experimenter presents a chip to a child, indicating that the child may hold it if he wishes, and asks, "can you find me one that is the same color as this one?", the child does not in general move the sample around attempting to match it with the various chips in the array. Instead he tends to glance at the sample and then look at the array, finally selecting a chip. Often he looks back at the sample, but seldom brings the trial chip next to the sample for a "same-or-different" judgement. This could be forced, perhaps by requesting the children to put the chip they select next to the sample, but there are two reasons for not doing this. The first is that we are interested in what the child considers to be "the same color," so we do not want to bias him away from his natural usage. The second reason is that by leaving the task open in this way, it is more similar to the recognition task, to be discussed below. In fact, the only difference between the matching and recognition tasks will be the five second delay between the presentation of the stimulus and the array. Therefore, the procedure in the matching task is to present a chip to the child and ask him to give the experimenter one "just like it."

Recognition Task

This is the usual recognition task, simplified for the younger subjects used in this experiment. The child is told that he will be shown a color, which will then be hidden, and that he must find a color just like it. The array is covered, a chip is presented and the child told to look at it carefully, the chip is removed, and five seconds later the array is uncovered.

Naming Task

In this task, the child is told that "now I want you to tell me about some colors." Single chips are presented and the child is asked "what do you call this color" or "do you know what is the name of this color?"

Fetching Task

All fourteen color chips are placed before the subject (not ordered in an array) and he is asked "can you give me all the _____ chips?" The three names most frequently given by the particular child in the naming task are thus given back to him. This serves both as a check on the consistency of his naming, and to illuminate the relationship of these two kinds of naming behavior.

Ordering of Stimuli within Tasks

Three tasks--matching, naming, and recognition--required an ordered presentation of the fourteen stimuli. Three orders are presented in Appendix A. In each of them, each successive pair of colors is separated by at least twenty-four steps of the Farnsworth-Munsell series (hence, four chips apart in the set of colors used in this experiment.) Each order occurred equally often matched with each of the tasks, and each child was given each of the orders. That is, one-third of the children had order I on the matching task, order II on the naming task, and order III on the recognition task; one-third had orders II, III, I, respectively; and one-third had orders III, I, II, respectively.

Ordering of Tasks

As the effect of verbal tags on nonverbal behavior is being investigated here, the verbal tasks were last within each session. The experiment was done in two sessions of about 15 minutes, separated by two or three days. This limit was determined largely by the span of attention of the children. The first session began with the matching task, which was followed by the first part of a very short experiment using pictures. In this experiment which lasted about two minutes the children were asked to point out the referents of nonsense syllables in the picture. Last was the naming task. The second session began with the recognition task, which was followed by the second part of the picture experiment, which was similar to the first. The session concluded with the fetching task.

Subjects

Twenty-four children between the ages of 3 years, 11 months, and five years, 1 month, from two nursery schools completed the experiment.

It was not always possible to complete the experiment successfully; some girls were too shy to respond, especially on the verbal tasks, while some boys were just too energetic to sit still for fifteen minutes. However, it was only necessary to work with twenty-nine children in order to obtain the twenty-four complete sets of data.

CHAPTER III - Experimental Results

Introduction

The data from the experiment are presented in Appendix B. The data for the matching, recognition, and naming tasks consist of one response to each color on each task by each child. They will be discussed at three levels. First is the analysis of performance on each task separately. The second level is the relation of performance on the various tasks, and the third is the nature of development on both of the above levels.

Matching Task

Table 1 presents the number of correct responses and the mean error for each child on the matching task, based on his fourteen matching responses. The mean absolute error considers only the magnitude of each error (the possible range is 0 to 7) and is a measure of the accuracy with which the child responded. The mean algebraic error includes the direction of error; a positive sign indicates an error in the direction of higher numbered colors and a negative sign indicates an error in the direction of lower numbered ones (the possible range is -6 to +7). The mean algebraic error is a measure of the central tendency of the child's responses, and will reflect any systematic bias. Purely random responding would produce a mean absolute error of 3.5 and a mean algebraic error of 0.

Figure 1 illustrates the relationship of matching accuracy to color. Table 2 includes the mean algebraic error by colors for the matching task,

TABLE 1

Number of Correct Responses and Mean Absolute and Algebraic Errors on Matching Task

Subject	Correct Responses	Mean Absolute Error	Mean Algebraic Error
1	7	.7	.3
2	6	.9	.0
3	6	.7	.1
4	6	1.1	.3
5	3	1.4	.1
6	8	.4	.0
7	8	.5	.2
8	5	.8	.1
9	8	.5	-.1
10	5	.8	-.1
11	2	1.1	.3
12	12	.1	-.1
13	13	.1	-.1
14	7	.7	.3
15	9	.4	.2
16	1	2.8	-.4
17	8	.6	.4
18	8	1.4	-.1
19	5	1.5	.9
20	6	.6	.1
21	4	1.6	-.4
22	7	.7	.6
23	4	.8	.1
24	4	1.1	.1
Mean	6.3	.85	.12

TABLE 2

Mean Algebraic Error by Color
for Matching and Recognition Tasks

Color	Mean Algebraic Matching Error	Mean Algebraic Recognition Error
1	-.67	-.54
2	.33	.63
3	.00	-.50
4	.54	1.83
5	1.04	.42
6	1.25	1.17
7	.79	.67
8	.04	.25
9	.38	-.25
10	-.21	-.54
11	-.63	-.79
12	-.25	-.63
13	-.17	-.08
14	-.75	-.29
Mean	.122	.095

as well as for the recognition task.

In this task, the colors can be characterized by their popularity as responses, i.e., how often each color chip was selected as a matching response to another color chip, summing across subjects. The colors vary greatly in their popularity, as indicated in Table 3.

TABLE 3

Popularity of Colors as Responses on Matching Task

Color	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Popularity	12	19	32	15	16	11	30	35	30	34	19	31	32	20

The Spearman rank correlation (used for all correlations in this dissertation) between mean absolute error and popularity is -.70 ($p<.01$). This suggests that the variation in accuracy with color shown in Figure 1 is a consequence of the variation in popularity, since there is a high probability that popular colors will be selected as matching responses to themselves.

Recognition Task

Table 4 presents the number of correct responses and the mean error for each child on the recognition task. Ten of the children spontaneously named each color as it was presented by the experimenter during the recognition task; they are starred in the table. Means were computed for all twenty-four children, for the ten namers, and for the fourteen non-namers. The difference between the namers and the non-namers in mean absolute error, while suggestive, was not significant.

TABLE 4

Number of Correct Responses and Mean Absolute and

Algebraic Errors on Recognition Task

Subject	Correct Responses	Mean Absolute Error	Mean Algebraic Error
1	3	2.9	-.4
2*	3	1.0	.3
3	4	1.1	.3
4	1	3.7	-.4
5*	4	1.1	-.3
6	10	.3	-.3
7	7	.6	-.1
8	2	1.6	-.7
9	7	.6	.0
10	4	1.3	.3
11*	5	.9	.1
12*	6	1.4	-.1
13	7	1.1	.1
14*	4	1.7	.9
15*	7	.6	.4
16	8	1.6	.9
17	3	.9	.4
18*	2	1.4	.1
19	2	2.4	-.1
20*	4	1.4	-.1
21	4	2.9	-.1
22	6	1.2	.6
23*	6	1.1	-.1
24*	4	1.2	.6
Mean	4.7	1.4	.10
Namers (*)	4.5	1.2	.2
Non-namers	4.8	1.6	.04

Figure 2 shows the relationship of recognition accuracy to color.

The popularity of colors as recognition responses is tabulated in Table 5.

TABLE 5

Popularity of Colors as Responses on Recognition Task

Color	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Popularity	16	19	36	14	26	10	24	25	35	44	19	20	30	18

The correlation between mean absolute error and popularity on the recognition task is .46 ($p<.05$).

Naming Task

Table 6 presents information on the naming performance of each child, based on his fourteen naming responses. The modal number of color names used by the children was four, but only four children used less than four names. This agrees well with the informal observation of Lenneberg (personal communication) that at least four terms are required for a stable system of color names.

The circular array of colors used in this experiment is divided up by each child into naming classes, i.e., sets of colors all of which were given the same name by him. Such a division into naming classes is consistent if each naming class consists of a set of adjacent colors. The consistency of naming is remarkable. Eighteen children were completely consistent in this sense, despite their often nonstandard use of color names, e.g., subject 19, who used only blue and white for the entire array. Of the remaining six children, only three had more than one deviation from such

TABLE 6

Summary of Color Naming by Each Child

Subject	Number of Color Names	Number of Colors Not Named	Average Size of Named Category
1	5	0	2.5
2	4*	3	2.8
3	5*	0	2.8
4	1*	9	5.0
5	4*	0	3.5
6	4	4	2.5
7	5*	0	2.8
8	5	0	2.8
9	3*	2	4.0
10	4*	0	3.5
11	4*	0	3.5
12	6	0	2.3
13	5*	0	2.8
14	4*	1	3.3
15	6*	0	2.3
16	4*	1	3.3
17	5*	0	2.8
18	4	0	3.5
19	2*	0	7.0
20	4*	0	3.5
21	3	1	4.3
22	5*	0	2.8
23	5*	0	2.8
24	4*	0	3.5
Mean	4.2	.9	3.3

* - consistent naming performance

consistency. Seven children did not name all the colors. In general however, those colors to which no name was given also were adjacent, e.g., subject 4. For the remainder of the analysis, those colors to which a child did not give a naming response will be taken as a naming class.

Blue and green were the most common terms used (nearly all the children used them). Brown, purple, and pink were the next most common. However, one should be cautious about inferences concerning the acquisition of these terms, as the color array used was highly specialized and was not a representative sample of the entire color space.

Table 7 analyzes the data by colors. For the later discussion, it will be very useful to have a measure of the codability of colors. Because each color was presented just once to each subject, and virtually all of the names elicited were single words, no intra-subject index can be computed. However, Brown and Lenneberg (1954) found that naming agreement, an inter-subject index, correlated highly with intra-subject measures. Codability was thus computed using their formula:

$$\text{Codability} = \begin{array}{l}\text{Number of time the most}\\ \text{popular name was given}\end{array} - \begin{array}{l}\text{Number of distinct}\\ \text{names given}\end{array} + 24,$$

where the final term is an additive constant to insure that the index will be positive. If all the subjects gave the same name to a color, the index would equal 24-1+24=47; while if each subject gave a different name, the index would equal 1-24+24=1.

This naming performance agrees well with that obtained by Lenneberg (1967) with adults, which is shown in Figure 3. The children used the names pink and purple in a fashion similar to the adult's use of rose and lavender, respectively.

TABLE 7

Summary of Names Assigned to Individual Colors

Color	Number of Names	Most Popular Name	Number of Times Given	Codability
1	7	pink	10	27
2	9	brown	15	30
3	7	brown	15	32
4	6	green	16	34
5	6	green	17	35
6	5	green	19	38
7	5	green	19	38
8	5	blue	12	31
9	5	blue	16	35
10	4	blue	20	40
11	6	blue	14	32
12	7	purple	13	30
13	7	purple	14	31
14	8	pink	9	25

Fetching Task

Again in the fetching task, nearly all of the children produced adjacent sets of colors as fetching responses to color names, even though the colors were presented to them for this task as an unordered set. Such fetching performance will be called consistent.

Relation of Performance on Matching and Naming Tasks

As was seen in Chapter I, the correlation between codability and accuracy for the colors is not a suitable measure for detecting the influence of verbal labels on nonverbal tasks. Instead, the number of matching responses which came from the same naming class as the stimulus color will be taken as such a measure. 271 of the 336 matching responses came from the same naming class as the stimulus color. This is to be expected, as most response colors are near to the stimulus color, and therefore are likely to be given the same name. To determine if there are significantly many more matching responses of this type than can be explained on the basis of the above considerations, a random model will be constructed for each child on the matching task. This model is defined by the set of probabilities:

$$p_{-6}, p_{-5}, \cdots , p_{-1}, p_0, p_1, \cdots , p_6, p_7,$$

where p_i is the probability of a response i steps to the right of the stimulus color. p_0 is the probability of a correct response, p_{-2} is the probability of a response two colors to the left of the stimulus, and so forth. These probabilities can be estimated from each child's actual responses, by "lumping" the colors together. The model, then, is based on each child's matching responses, but does not take into consideration his naming behavior.

Next we can compute the expected number of times this model will produce as a matching response a color which came from the same naming class as the stimulus. This is done by summing the probabilities of responding with any of the other colors in the naming class of the stimulus, and iterating for each of the fourteen stimuli. Formally,

$$\mu = \sum_{i=1}^{14} \text{Pr [Producing Response in Same Naming Class as Color i]}$$

$$= \sum_{i=1}^{14} \sum_{j=1}^{14} P_{j-i} \; N_{ij}, \quad \text{where } N_{ij} = 1 \text{ if color i was given the}$$

same name as color j, 0 otherwise.

These predicted numbers, one for each child, are presented in the first column of Table 8. They represent the expected number of matching responses coming from the same naming class as the stimulus if only the processes mentioned earlier--the likelihood of a nearby color as a matching response, and the likelihood that a nearby color would be given the same name--were operating. In the second column of the table are presented the actual number of such responses. The difference between the columns can be tested by the sign test, and it can be concluded at the .001 level that there are more responses coming from the same naming class as the stimulus than can be accounted for by the random model.

In contrast to this analysis, which displays a clear relation between the names given to colors, and the way the colors are manipulated in the matching task, the correlation between matching accuracy (the inverse of matching absolute error) and codability is $-.46$ ($p < .05$), in apparently the "wrong" direction. That is, the more codable the color, the less accurate the responses to it; just as Burnham and Clark (1955) found with a similar color array. This is further evidence of

TABLE 8

Predicted Versus Observed Number of Response Colors Falling in Same Naming Category as Stimulus on Matching Task

Subject	Predicted	Observed
1	10.43	11
2	10.14	12
3	10.57	11
4	11.71	14
5	8.29	10
6	11.43	13
7	11.64	13
8	6.21	8
9	12.00	12
10	10.86	14
11	9.64	9
12	13.00	14
13	13.64	14
14	10.71	11
15	11.86	13
16	4.86	5
17	11.14	12
18	10.57	11
19	11.07	12
20	11.71	12
21	6.93	6
22	10.43	12
23	10.14	13
24	9.86	9

Observed > Predicted
$p < .001$
(sign test)

the unsuitability of codability in experiments such as the present one.

Another way to view this phenomenon is shown in Figure 4. If matching responses coming from a different naming class than the stimulus color are less frequent than would otherwise be expected (as is shown in Table 8), this should show up most visibly at the ends of naming classes. At the boundary of such classes, an adjacent color, which would normally be expected to be a frequent matching response to the given color, has a different name, according to the child. Such a color is not likely to be selected, and this should result in a bias in the central tendency of his responses, a bias in the direction of the center of the naming class. In particular, at the left end of naming classes, (the end with the lowest number), the mean algebraic error should be positive, whereas at the right end, the mean algebraic error should be negative. The figure summarizes the errors for the 91 naming classes of size two or larger. There is some overall bias in the direction of error, but the effect just mentioned, the tendency for matching errors to be made in the direction of the center of the naming class, is apparent.

If the frequency of correct matching responses is considered, the values at the left and right ends of naming classes (.47 and .43) do not differ appreciably from the overall frequency of correct responses (.45). Similarly, the accuracy (as defined by mean absolute error) at the ends of naming classes (.80 and .78) is not very different from the overall accuracy (.85). It does not, therefore, appear to be the case that the entire distribution of matching responses is shifted at naming class boundaries, for in that case the number of correct responses and the accuracy would be diminished. Instead, the phenomenon seems to be much simpler: responses outside the naming class are inhibited.

Relation of Performance on Recognition and Naming Tasks

As the data from the recognition task are of the same form as that from the matching task, i.e., one recognition response to each of the fourteen stimulus colors for each of the subjects, the same type of analysis can be used. Again, the number of responses which came from the same naming class as the stimulus color will be used as a measure. A random model can be constructed, this time using the child's error frequencies on the recognition task, and a predicted number of such responses for each child are shown in Table 9. Again the random model cannot account for the number of responses coming from the same naming class as the stimulus ($p < .05$).

Contrary to the expectation, the effect here is weaker than in the matching task. If the spontaneous namers and non-namers are considered separately, the situation is clarified. Of the ten namers, nine have differences in the direction of more responses from the same naming class as the stimulus than predicted ($p < .015$); whereas of the fourteen non-namers, only seven have differences in that direction (N.S.). Thus, only the spontaneous namers show a significant tendency to produce such responses.

The correlation between codability and recognition accuracy (the inverse of recognition absolute error) is .03 (N.S.).

Figure 5 presents the mean algebraic error overall and at the ends of naming classes. Despite the overall bias, the tendency for errors to be made in the direction of the center of naming classes is apparent. Just as was the case for the matching task, the frequencies of correct recognition responses at the left and right end of naming classes (.27 and .33, respectively) do not differ greatly from the overall frequency

TABLE 9

Predicted Versus Observed Number of Response Colors Falling in Same Naming Category as Stimulus on Recognition Task

Subject	Predicted	Observed
1	4.78	6
2*	9.36	10
3	9.21	8
4	7.14	10
5*	9.43	11
6	12.29	11
7	11.14	12
8	3.64	2
9	11.71	13
10	9.29	12
11*	10.36	13
12*	8.21	10
13	9.00	9
14*	8.07	8
15*	10.71	11
16	9.43	8
17	9.36	10
18*	6.43	8
19	9.21	12
20*	9.00	10
21	6.36	5
22	9.29	9
23*	10.07	12
24*	9.29	10

* Namers

Observed > Predicted
Overall $p < .05$
Namers $p < .015$
Non-namers N.S.

(.34). Similarly, the mean absolute errors at naming class boundaries (1.64 and 1.36) are the same as the overall value (1.42). Since the number of correct responses and the accuracy are not diminished at the naming class boundaries, the phenomenon appears to be the same as was observed in the matching task: responses outside the naming class are inhibited.

Relation of Performance on Matching and Recognition Tasks

The subject's performance on the matching and recognition tasks can be compared in several ways. The mean absolute error on the two tasks, as shown in Figures 1 and 2 have a correlation of .65 ($p < .01$). The popularity of the colors as responses in the two tasks, as shown in Tables 3 and 5, have a correlation of .70 ($p < .01$).

For a more revealing comparison, it is necessary to split the notion of accuracy as defined by mean absolute error into two parts. The first component is the central tendency of each child's errors. This is measured by his mean algebraic error, which reflects any systematic bias in his responding. For example, subject 14 in the recognition task on the average responded with a color almost one full step to the right of the stimulus color. The second component is the subject's variability about this mean, which is measured by the standard deviation of his errors about the mean. This is an index of the "spread" of his responses.

The correlation between the mean algebraic error for each subject on the two tasks (as tabulated in Tables 1 and 4) is .02 (N.S.): this rules out any systematic, individual bias in their responding.

The standard deviation of response errors about the mean algebraic error for the matching and recognition task for each subject is shown in Table 10. In general, the spread of responses, as measured in this way, is greater for the recognition task than for the matching task. This reflects the subjects' generally poorer performance on the recognition task.

The most direct measurement of the similarity in responding on the matching and recognition tasks is the mean absolute difference between corresponding responses to the fourteen colors. These are also presented in Table 10. They are in most cases, less than the standard deviation of the response error on the matching task; that is, there is no significant shift in responding from one task to the other.

The importance of the measures introduced in this section lies in their application to the namers and non-namers separately. This will be done in the next chapter, in the context of the model introduced there.

Relation of Performance on Naming and Fetching Tasks

Eighteen of the children produced adjacent sets of colors as responses to color names on the fetching task, even though the colors were presented in an unordered way for this task. Sixteen children performed consistently on both the naming and fetching tasks. Their performance on the two tasks is compared in Table 11. It can be seen that in the majority of cases the categories corresponding to a given color name on the two tasks are either identical or one (properly) includes the other. There is no apparent difference in the sizes of categories produced in the two tasks.

TABLE 10

Difference Between Matching and Recognition Responses to Same Stimulus and Standard Deviation of Response Error

Subject	Mean Absolute Difference between Matching and Recognition Responses to Same Stimulus	Standard Deviation of Response Error on Matching Task	Standard Deviation of Response Error on Recognition Task
1	1.80	1.22	3.61
2*	.49	1.26	1.22
3	.59	1.05	1.70
4	2.32	1.79	4.27
5*	.84	1.73	1.38
6	.41	.63	.45
7	.49	.77	1.05
8	.88	1.05	1.97
9	.20	.77	.84
10	.71	1.05	1.79
11*	.80	1.41	1.26
12*	.80	.32	2.19
13	.63	.32	1.70
14*	.84	1.10	2.47
15*	.29	.63	.71
16	1.96	3.42	2.63
17	.29	.95	1.10
18*	.67	.63	2.02
19	2.12	2.32	3.26
20*	.76	.77	2.09
21	1.88	2.19	3.91
22	.63	.84	1.92
23*	.59	.95	1.84
24*	.59	1.45	1.45
Mean	.90		
Namers (*)	.67		
Non-namers	1.07		

TABLE 11

Comparison of Performance on Naming and Fetching Tasks by Consistent Namers and Fetchers

Subject	Color Name	Fetching Category	Relation	Naming Category
2	green	7–9	⊂	4–9
	brown	1–3	⊃	2–3
	blue	9–11	⊃	10--11
3	green	4–7	=	4–7
	brown	1–3	=	1–3
	purple	11–13	=	11 13
5	green	5–8	~	4–7
	blue	9–11	⊂	8–11
	pink	13–1	~	13–1
7	green	4–10	⊃	4–9
	purple	12–1	~	11–14
	brown	2–3	=	2–3
9	blue	9–11	⊂	8–11
	green	5–7	⊂	4–7
	pink	13–1	⊂	12–1
10	tan	3–4	⊂	2–5
	blue	9–12	=	9–12
	green	6–8	=	6–8
13	blue	9–12	⊃	9–11
	brown	14–3	⊃	2–3
	green	5–8	~	4–8
14	purple	7–13	>	11–14
	green	7–10	~	4–7
	blue	6–9	~	8–10

TABLE 11- Continued

Subject	Color Names	Fetching Category	Relation	Naming Category
15	pink	14	C	12-14
	green	4-7	=	4-7
	beige	8 9	=	8 9
16	blue	8-11	~	7-10
	green	5-7	Ɔ	5-6
	purple	11-13	C	11-14
17	green	6-7	C	4-7
	blue	9-11	C	8-11
	red	13-14	~	14-1
19	blue	5-12	Ɔ	7-12
	white	12-6	Ɔ	13-6
20	purple	13-14	C	11-14
	green	5-6	C	4-7
	blue	7-10	Ɔ	8-10
22	blue	9-10	C	8-11
	green	5-9	>	4-7
	brown	3	~	2
23	purple	12-1	Ɔ	12-14
	green	5-9	C	4-9
	brown	2-3	=	2-3
24	blue	9-11	C	8-11
	green	4-8	Ɔ	4-7
	purple	12-14	C	12-1

= - equality
Ɔ - fetching category includes naming
C - naming category includes fetching
~ - same size, but not identical
> - fetching category larger

TABLE 11- Continued

Subject	Color Name	Fetching Category	Relation	Naming Category
Mean size	blue	3.8		3.8
	green	3.9		4.1
	brown	2.5		2.2
	purple	3.7		3.7
	pink	2.3		3.7

Relations: 9 =, 16⊂, 11⊃, 9~, 2 >.

Development of Performance on Color Tasks

For the purpose of studying the nature of development, the twenty-four children were divided into three groups of eight each, as is indicated in the data in Appendix B. The mean age of the first group was approximately 4;2, of the second group, 4;5, and of the third group 4;10.

The general nature of development is shown in Figure 6. As one would expect, there is a steady increase in matching and recognition accuracy, and an increase in the number of color names used. For matching and recognition performance, the variation with age is significant ($p < .01$ and $p < .02$, respectively; Kruskal-Wallis one-way analysis of variance).

As a crude measure of the influence of the verbal labels on the matching task, the average excess of the observed over the predicted number of matching responses coming from the same naming class as the stimulus (as defined on page 28) can be used. Table 12 shows the relatively constant level of this measure.

TABLE 12

Average Excess of Observed over Predicted Values of Matching Responses Coming from the same Naming Class as the Stimulus by Age Group

Age Group	I	II	III
Average Excess	.42	1.30	.99

If the nature of development on the matching task were simply toward increased accuracy, one would expect this measure to decrease, since the influence of the names must weaken as discrimination within naming categories increases. The relative constancy of the measure suggests that, on the contrary, the increased accuracy on the matching task is a result of more highly developed (in particular more numerous) naming categories. A mechanism that inhibits matching responses outside the naming category of the stimulus would be increasingly effective with finer naming categories.

Table 13 presents the average excess of observed over predicted number of recognition responses coming from the same naming class as the stimulus for the recognition task. There appears to be a definite increasing tendency, but it is not significant.

TABLE 13

Average Excess of Observed over Predicted Values of Recognition Responses Coming from the Same Naming Class as the Stimulus by Age Group

Age Group	I	II	III
Average Excess	.19	.93	1.04

If this measure is actually increasing, performance on recognition may be improving both because of the development of finer naming categories and because of the increased inhibition of responses outside the naming categories of the stimulus.

CHAPTER IV - Discussion

The Influence of Verbal Labels on Nonverbal Tasks

The experimental results presented in the preceding chapter demonstrate clearly a relation between the name given to a color, and the way the color is manipulated in the matching and recognition tasks. The large number of matching and recognition responses that came from the same naming class as the stimulus cannot be accounted for by a random model that is based on the child's performance on the matching or recognition task but does not take into account this naming behavior.

This phenomenon can be stated in terms of predicting the responses on one task from knowledge of the responses on another. As an example, consider the performance of subject 7 on the matching and naming tasks. This subject's matching response probabilities (as defined on page 27) are:

$$p_{-1} = .14,\ p_0 = .57,\ p_1 = .21,\ p_2 = .07,\ p_i = 0 \text{ for all other } i$$

This set of probabilities defines a distribution of probabilities of particular responses, and can be considered as a predictor of the child's responses. The mean absolute difference between this predicted distribution and the actual matching responses is an index of predictive accuracy, and can be computed with the following formula:

$$D = \sum_{i=1}^{14} \frac{1}{14} \sum_{j=1}^{14} [\text{Pr (Response j)} \cdot \text{Error of Response j}]$$

$$= \sum_{i=1}^{14} \frac{1}{14} \sum_{j=1}^{14} p_{j-i} \left| j - m_i \right| \text{, where } m_i \text{ is the actual matching}$$

response to stimulus color i. For subject 7, D equals .77.

This index has been computed without reference to subject 7's naming behavior. This subject made thirteen out of fourteen matching responses from the same naming class as the stimulus. Now let us add to the set of probabilities above the condition that responses from outside the naming class of the stimulus are <u>never</u> made (which is slightly stronger than the fact) and specification of his naming classes as:

(1) (2,3) (4,5,6,7,8,9,) (10), (11,12,13, 14).

The predictive accuracy of this new distribution can be computed with the following formula:

$$D' = \sum_{i-1}^{14} \frac{1}{14} \left[\sum_{j=1}^{14} \frac{1}{\sum_{K=1}^{14} N_{ik} P_{k-i}} P_j N_{ij} \left| j-m_i \right| \right], \quad \text{where } N_{ij} = 1$$

if colors i and j were given the same name, 0 otherwise. The denominator of the fraction in the expression compensates for the elimination of the possibility of responses outside the naming class of the stimulus, in order to maintain the sum of the (remaining) response probabilities equal to 1.0. For subject 7, D' equals .63. Thus predictive accuracy can be increased through the use of information about the subject's naming.

In the opposite direction, predictions about naming can be made on the basis of matching responses. In particular, if color i was given as a matching response to color j, it can be predicted that the two colors will be given the same name. It must also be assumed that naming categories will be continuous segments. In general, these predictions will be correct, as only responses outside the naming class of the stimulus, which are infrequent, will lead to false predictions. However, the predictions will be weak; that is, they will imply a large number of small classes. This is because performance on the matching task by the children was quite good, and it will therefore seldom be possible to claim that two colors which are far apart will have the same name, since errors of that size do not occur. In the case of subject 7, the following classes can be predicted and compared with the actual classes:

Predicted:	(1)	(2,3)	(4,5)	(6,7,8)	(9,10)	(11)	(12)	(13)	(14)
Actual:	RE	BR BR	GR GR	GR GR GR	GR BL	PU	PU	PU	PU

Note that just one response from outside the naming class of the stimulus was made by this subject, namely, a matching response of color 10 to stimulus color 9.

This phenomenon is not an example of the general phenomenon of categorial perception. As discussed in the first chapter, the distinguishing characteristic of categorial perception is enhanced discrimination at category boundaries relative to discrimination within categories. Two measures of discriminability--frequency of correct response, and mean absolute error--do not vary in this way; in particular, the measures are essentially constant, and this holds for both matching and recognition. The opposite mechanism, which was termed attractive perception in the first chapter, is also not present, for it leads to the prediction that performance at boundaries will be inferior to that within categories; this is contradicted by the same evidence.

Instead, the mechanism seems to be much simpler. Responses from outside the naming class of the stimulus color (which will be called <u>extra-categorial</u> responses) are inhibited. Concurrently, there is a tendency for certain colors to be highly popular as responses, both to themselves--which leads to high accuracy for these colors--and to other colors.

The underlying nature of this popularity effect is problematic. One hypothesis is that there are "best instances" of the various color names used by the children, and that these colors were most likely to be selected whenever the stimulus color belonged to that naming class. From this it would follow that codability and popularity should be positively correlated. But for matching, the correlation is only .03 (N.S.), and for recognition, .33 (N.S.). Nevertheless, the possibility remains that this process does occur in each child, but that the particular best instances vary from child to child so greatly that the inter-subject index of codability, which serves to identify the best instances, is not valid. With the data available, it is not possible to investigate this further, as there is no way to tentatively identify the best instances except by popularity, and this renders circular the correlation between "best-instance-hood" and popularity, which is the only means of confirming this hypothesis. This could have been circumvented if the children had been asked to name the colors twice, in order that an intra-subject codability index could be computed.

Another hypothesis concerning the variation in popularity is that the colors simply differ in their attractiveness to children, and thus introduce a response bias. Although this agrees with the experimenter's intuition, it is not testable with the present data.

This inhibition of extra-categorial responses is stronger in the matching task than in the recognition task. On the basis of the weakest form of the Whorf hypothesis, that the effect of language is essentially limited to memory, the opposite result would be expected.

This suggests the following dual mechanism for colors. Colors are entered into storage in two different forms. One is representational, a function of the physical properties of the stimulus. The other is symbolic; i.e., some identification of the category (naming class) to which the stimulus color belongs. In general, both forms are produced, even for very short-term storage as in the matching task. But in young children such as those in the experiment, the symbolic storage is very short lived. If the category identification is present at the time of response selection, it will influence the selection of a response by inhibiting extra-categorial responses. This is precisely what happens in the matching task. But if it is no longer available (for whatever reason), the response will be more directly controlled by the representational encoding. This is what happens, in general, in the recognition task. But the spontaneous namers, having rehearsed the name, are much more likely to have the symbolic form available to them when it is time to select a response color, and hence they show the inhibition. The remainder of this section will be devoted to a discussion of this model.

According to this model, the matching and recognition responses of the namers should be more similar than the matching and recognition responses of the non-namers. This follows from the statement that for the namers, the recognition responses are selected using both the representational and symbolic encoding, just as for the matching task. But the non-namers can use only the representational encoding on the recognition task instead of using both, as they did in the matching task. Table 10 shows that the mean absolute difference between matching and recognition responses to the same stimulus was .67 for the namers and 1.07 for the non-namers ($p < .01$; Mann-Whitney U Test).

For the non-namers, there is a clear tendency for the response errors to be more spread out in the recognition task, as indicated by a larger standard deviation of response error on the recognition task than on the matching task ($p < .01$; sign test). But this is not the case for the namers; the spread of their responses does not increase significantly. This is interesting, because their accuracy, as measured by mean absolute error, is not significantly better (Table 4). In fact, the frequency of correct responses is slightly lower for the namers. The most plausible explanation for this is that the distribution of responses based on the representational storage alone tends to be strongly unimodal, with the mode at the correct response, but with relatively long tails. The distribution of responses based on both forms of encoding loses the long tails (inhibition of extra-categorial responses), but the frequency of selection of the correct color is slightly depressed, as all the colors in the naming class are more likely to be confused than would be the case if the category identification were not present.

The fact that some of the children spontaneously named colors, and then had lower errors, though not significantly so, suggests that this experiment be interpreted in terms of mediation. Recently, Flavell and his associates (Flavell, Beach, and Chinsky, 1966) have been investigating mediation, and the lack of mediation, in children in some detail. The hypothesis that has been generally held is the "mediational-deficiency hypothesis:" there is a stage in ontogenesis during which the child tends not to mediate or regulate his overt behavior verbally, despite the fact that he is able to understand and correctly use the words in question (Reese, 1962). Flavell et al. (1966) distinguish two distinct factors for explaining this often-reported inability of children to utilize their language skills for cognitive tasks: 1) a production deficiency, in which the child lacks an ability or disposition to produce would-be verbal mediators; and 2) a true mediational deficiency, in which the child produces the verbalizations but they fail to mediate his response. They showed, using a serial learning task with pictures, that kindergarten children were prone to the former deficiency, as evidenced by their failure to rehearse stimulus names. In a subsequent study (Keeney, Canizzo, and Flavell, 1967) two groups of first grade children were identified: those who spontaneously rehearsed (producers), and those who did not rehearse (nonrehearsers). The serial recall of the nonrehearsers was significantly poorer than that of the rehearsers. But when the nonproducers were induced by the experimenter to rehearse the items to be recalled, their level of recall rose to that of the spontaneous producers. When subsequently given the option of rehearsing or not, they tended to drop rehearsing, and their recall fell to the former level. Note that here the children not only were able to use the verbal label correctly; they could even utilize the labels for memory when induced to produce them. But in a free situation they failed to produce the mediator. Thus they manifested production deficiency but not mediational deficiency.

In the color experiment reported here, mediated responses can only be observed statistically, with the measure introduced on page 28. Any response color coming from the same naming class as the stimulus is potentially a mediated response, but these are frequent even without the operation of the inhibition effect discussed above. The responses of a subject can only be considered mediated if there are more of these responses from the same naming class as the stimulus than can otherwise be accounted for.

With this qualification of the notion of a mediated response, the non-naming children fail to produce the color name, and do not produce mediated responses. They appear to manifest production deficiency, in contrast to the namers, who produce the mediator and produce mediated responses. However, this interpretation is contradicted by the high degree of mediation shown in the matching task, where none of the children spontaneously named the colors. The argument is as follows: To explain the high degree of mediation in the matching task, it is necessary to postulate internal production of a category identification; this identification, whatever its form, mediates the responses. But if such an internal mediator is produced by all the subjects in the matching task, it must be produced by all the subjects, including the non-namers, in the recognition task. If this is the case, these non-namers show a mediational deficiency, rather than a production deficiency.

In this way, the subjects in the recognition task divide into two groups: those who produce a mediator internally and show a mediational deficiency, and those who overtly name and show mediated responses. There are two explanations for this division. First, the former group has the mediator available at the time of selection of response, but fails to use it. This is implausible in view of their successful utilization of the mediator in the matching task. Second, the internal producers do not retain the mediator in memory until response selection, in contrast to the overt namers, who do retain it. In this explanation, the crucial difference between the namers and the non-namers is that the former _rehearse_ the mediator, hence they are more likely to remember it. This is consistent with the general role of rehearsal in short-term memory.

Adult subjects are able to rehearse internally; it is not necessary to produce the items being remembered overtly. This does not seem to be possible for the children in this experiment. If they were able to do so, the non-namers (overtly) would also show mediated responses.

This hypothesis (mediational deficiency resulting from lack of memory for the mediator) could be further investigated by instructing the children to name each color as it was presented in the recognition task. The model proposed above leads to prediction of an increased degree of mediation, since the children do not lack the ability to utilize a mediator, but rather fail to remember the mediator. Thus instructing them to rehearse the name increases the probability that it will be available at the time of response selection. If this were done on the matching task also, no significant change would be expected, as the children are already producing the mediator internally, and do not have any difficulty remembering it for the very short interval required in the matching task.

How does this model relate to the questions raised in the first chapter? It demonstrates, first, that attributes are categorized even in simple perceptual tasks, and that this is true even at an early stage of development, when the categories themselves are not yet fully developed.

Perhaps the weak form of the Whorf hypothesis, the form that states that production of verbal labels influence how we see the world, is based on a false assumption: that the act of seeing the world is a unitary act. What this model suggests is that visual experience of this type is stored in two essentially different ways: a representational storage, and a category identification. Both of these are always produced, even in tasks as simple as matching. The categorization cannot be short-circuited. As long as the categorization is available, it will influence the output, whether or not the output is verbal. Nevertheless, these two encodings are distinct, and each can have a different internal history. This can be seen in the distinction between the namers and non-namers in the recognition task; the latter simply lost the category identification while preserving the representational storage. The fact that the namers did not do much better on the recognition task than the non-namers suggests that the representational storage held up just as well in both groups, despite the great difference in mediation.

It is in the postulation of two distinct, but concurrent encodings that this model differs from the weak form of the hypothesis, in which the category to which the stimulus belongs biases the representational storage, which is essentially unitary.

The Relation of the Naming and Fetching Tasks

A secondary question of interest in this study was the relation between the naming and fetching tasks. The discussion in the first chapter suggested that a significant difference in the size of the classes of colors corresponding to a given color name might imply the reducibility of one task to the other. The lack of a significant difference in these sizes, as shown in Table 11, means that neither asymmetrical theory of their relation can be verified. Instead, naming and fetching must both be based on more fundamental processes.

The Development of Color Categories

The results of the experiment show an increasing accuracy on both the matching and recognition tasks, and an increasing number of color names with age. This developmental trend will be examined more closely in this section.

No such trend was found for the inhibition of extra-categorial responses. For the matching task, the fraction of subjects showing an excess of observed over predicted number of responses coming from the same naming class as the stimulus in each of the age groups was : (I) 7/8, (II) 7/8, (III) 6/8. Table 10 shows that the average excess of observed over predicted responses of this type remains relatively constant. This is an significant point. If improvement on this task were simply a matter of increasing accuracy, having nothing to do with the color names and their corresponding categories, this measure should decrease, as can be seen from the following argument. "Increasing accuracy" means a change in the p_i values, such that the mean error

$$\mu = \sum_{i=-6}^{7} |i| \, p_i$$

decreases. As this is carried to the limit, p_o, the probability of a correct response, tends to 1.0. As this occurs, the predicted probability of a response assigned the same name as the stimulus tends to 1.0 (each color is assigned the same name as itself), and the expected number of such responses tends to fourteen. At the same time, as the probability of a correct response tends to 1.0, we _must_ observe the actual number of such responses tending to fourteen. Thus the difference between the predicted and observed values tends to zero. But in general, it has been seen that this difference is greater than zero under the conditions of the present experiment. Thus the difference would decrease absolutely if the p_i values were changing in the way mentioned above.

Thus, the fact that it does not decrease, while the overall accuracy does increase, suggests another process. If a child is equipped with a mechanism which inhibits extra-categorial responses in some relatively constant fashion, it will be increasingly effective as the number of categories increases, for the size of each category will be smaller. And indeed, the average size of the named categories, i.e., excluding the "no name" category, for the three age groups was: (I) 3.8, (II) 3.2, (III) 3.0.

Each category boundary, then, can be considered as an additional discrimination made by the child. Only a longitudinal study, one in which this experiment would be administered to each child several times over a period of time, could determine the pattern of development of such discriminations. The simplest hypothesis is that each such discrimination is simply added to the previous system. This is another way of stating that color categories develop through differentiation; i.e., each new category is properly included in one earlier category. This hypothesis is probably too simple.

This leads to the second point to be made in this section. Although the naming behavior of a subject leads to a simple partition of the color space, i.e., a division into a set of mutually disjoint and exhaustive subsets, in fact the situation is more complex. Perfectly rectangular graphs are not obtained in Figure 3. Some colors are more appropriately described by a given color name than others, and presumably there is a best instance of each color name in the color space taken as a whole, and in any subspace of it, such as the set of stimulus colors used in this experiment. This fact, of which all adult language users are aware, suggests a mechanism for the development of color categories.

Perhaps the child acquires color names by associating each such term with a single best instance of it. Strictly speaking, this is impossible, but we can imagine a very small range of colors initially associated with the name. The categories are then enlarged by some kind of generalization. The generalization might be carried too far; then a new color name would be associated with some color in an old category, which would have to be split off and generalized into a new category. In this case, the category identification is equivalent to the best instance in the category.

However, the evidence from this experiment does not support such a theory. There is no detectable tendency of the children to work with best instances of color names. "Not working with best instances" means that the category identification is equivalent to the entire category.

The opposite hypothesis, that children first develop a complete system of categories, and only then learn that some colors are more appropriately described by color names than others, is too strong to be supported by this experiment. Perhaps only after a particular category has been established for some time can the child learn which colors fall most clearly into it, and this process must be repeated for each category. Again, a longitudinal study would be very revealing.

The third point is concerned with the act of naming the colors in the recognition task, and the role of rehearsal. For the recognition task, the fraction of subjects showing mediated responses (in the statistical sense) was: (I) 4/8, (II) 6/8, (III) 6/8. Table 13 shows an increase (not significant) in the average excess of observed over predicted responses coming from the same naming class as the stimulus. The fact that these measures do not decline is further evidence that what is happening here is not simply an increase in accuracy of memory (by the argument at the beginning of this section). Note that the developmental trend is not the result of an increasing number of spontaneous namers in this task; for the namers are distributed among the age groups as follows: (I) 4, (II) 4, (II) 3.

This experiment is concerned with a noncommunicative use of language. Vygotsky (1962) claimed that such nonsocial speech gradually became internalized in childhood to become inner speech. The child must learn to produce inner speech internally. It would not be too great a distortion of Vygotsky's views to say that for him, the child suffers from production deficiency. The results of this experiment suggest a different interpretation. All the children appear to be producing the mediator, a crude kind of inner speech. The problem is that some do not remember it, and this is mediational deficiency. The only mechanism they have to help them remember is rehearsal; and for children of this age, rehearsal means overt rehearsal. As the children become older, they acquire the ability to rehearse internally, and we would expect overt naming to decrease. Thus what becomes internalized may not be inner speech, which is internal from early childhood, but rather the ability to pay attention to it, and in particular, to remember it.

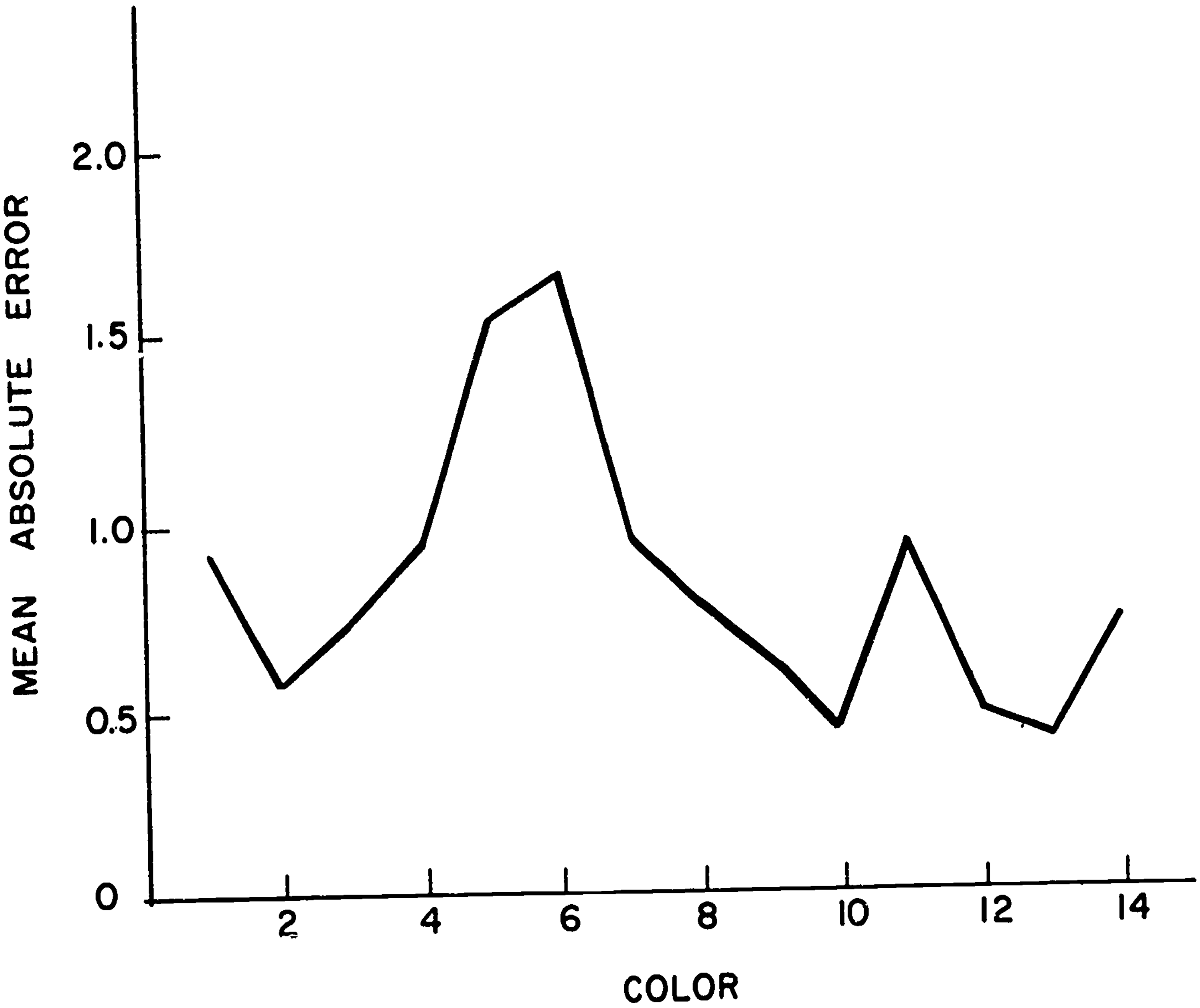

Figure 1. Mean Absolute Error on Matching Task by Color.

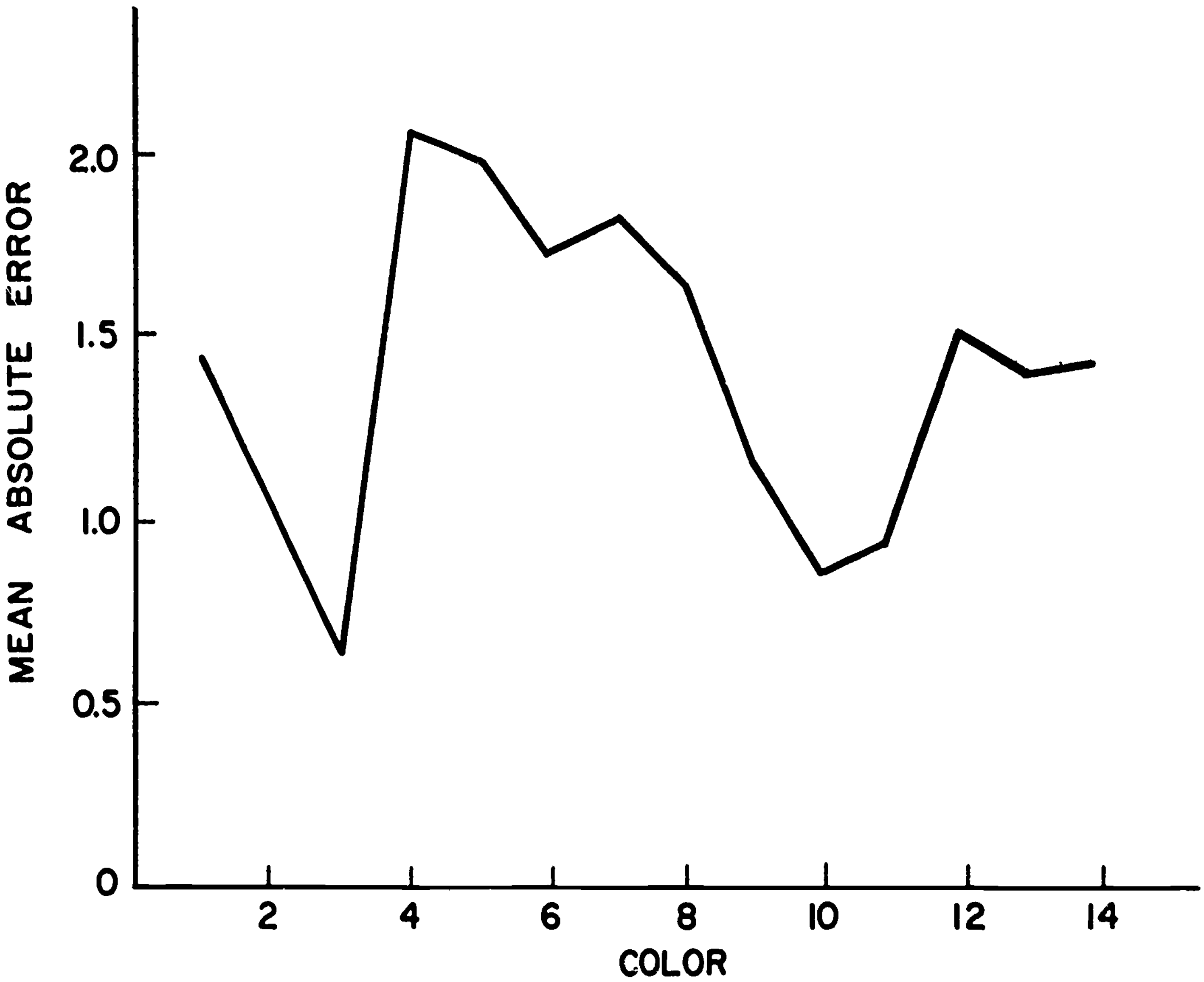

Figure 2. Mean Absolute Error on Recognition Task by Color.

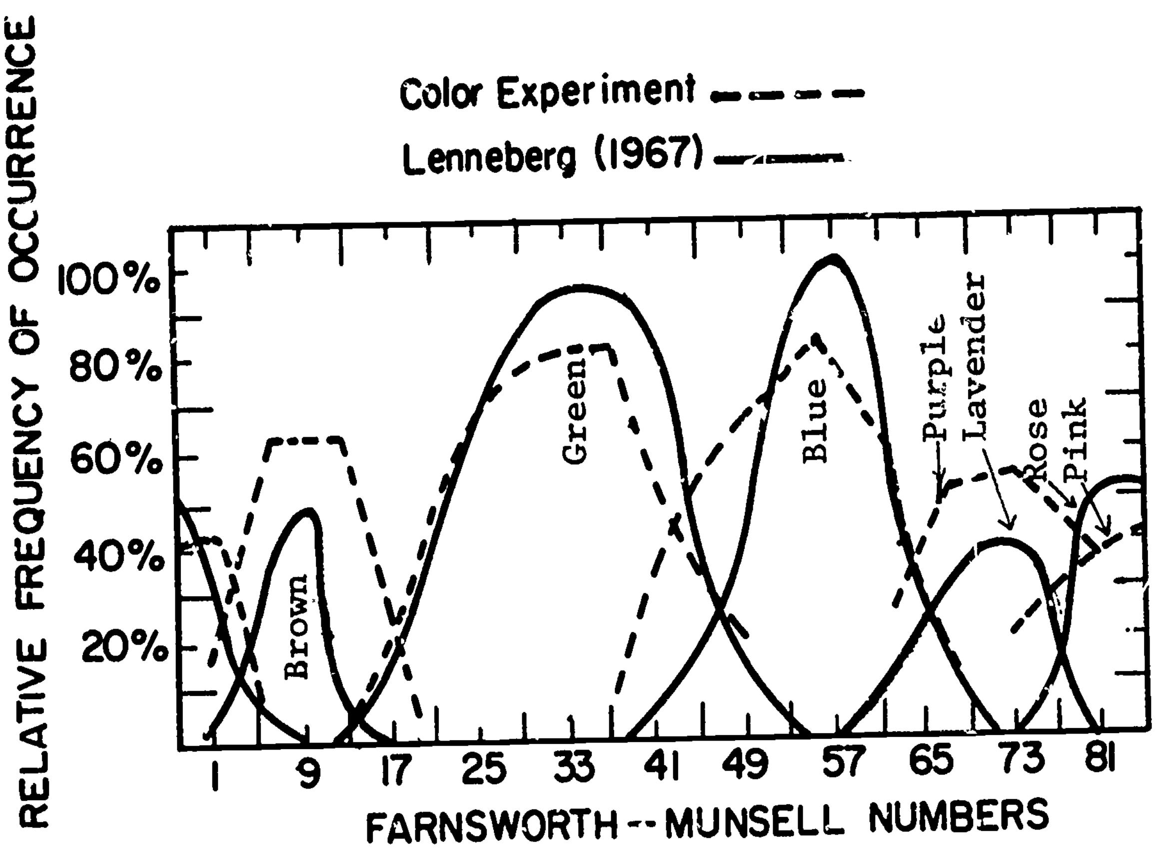

Figure 3. Distribution of Use of Color Names in Color Experiment and Comparison with Lenneberg (1967).

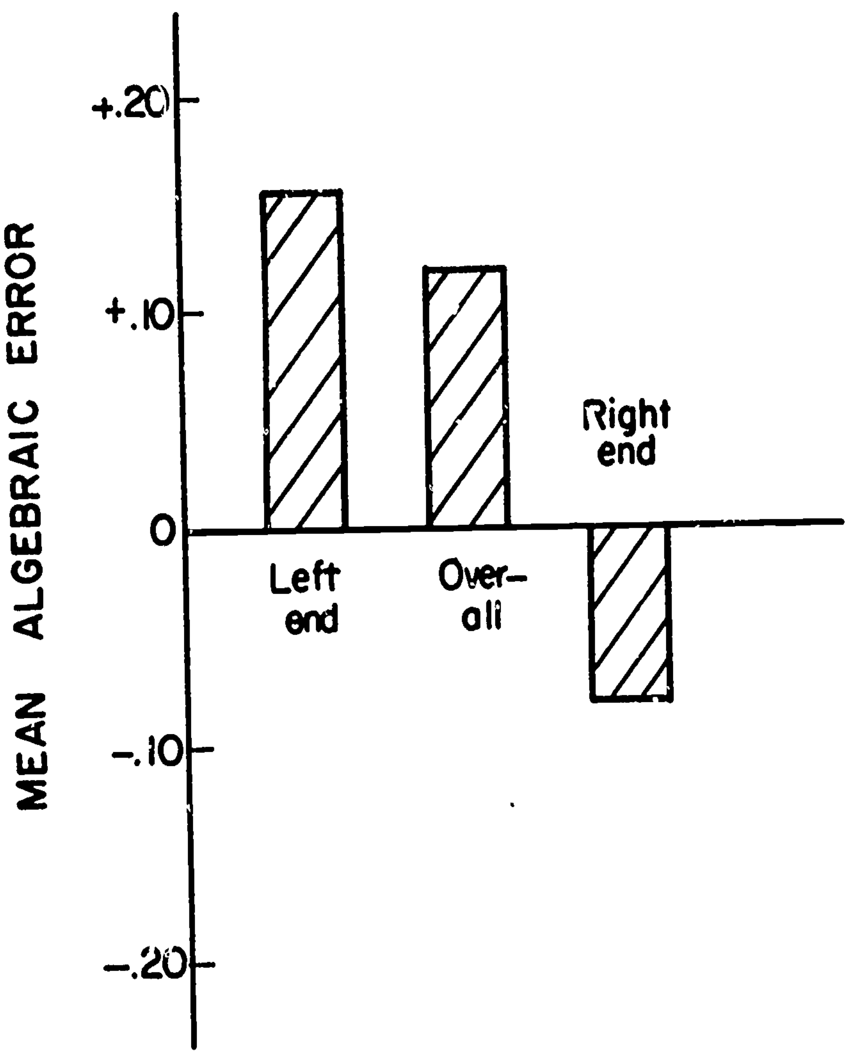

Figure 4. Mean Algebraic Error on Matching Task at Naming Class Boundaries.

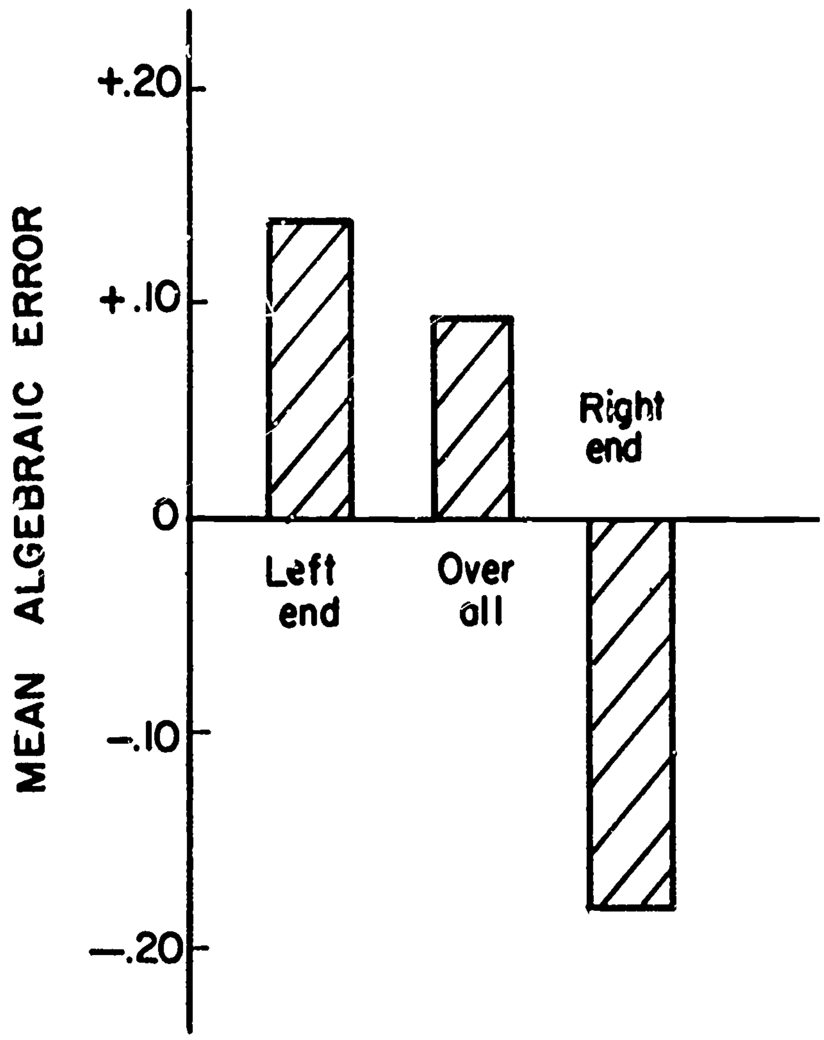

Figure 5. Mean Algebraic Error on Recognition Task at Naming Class Boundaries.

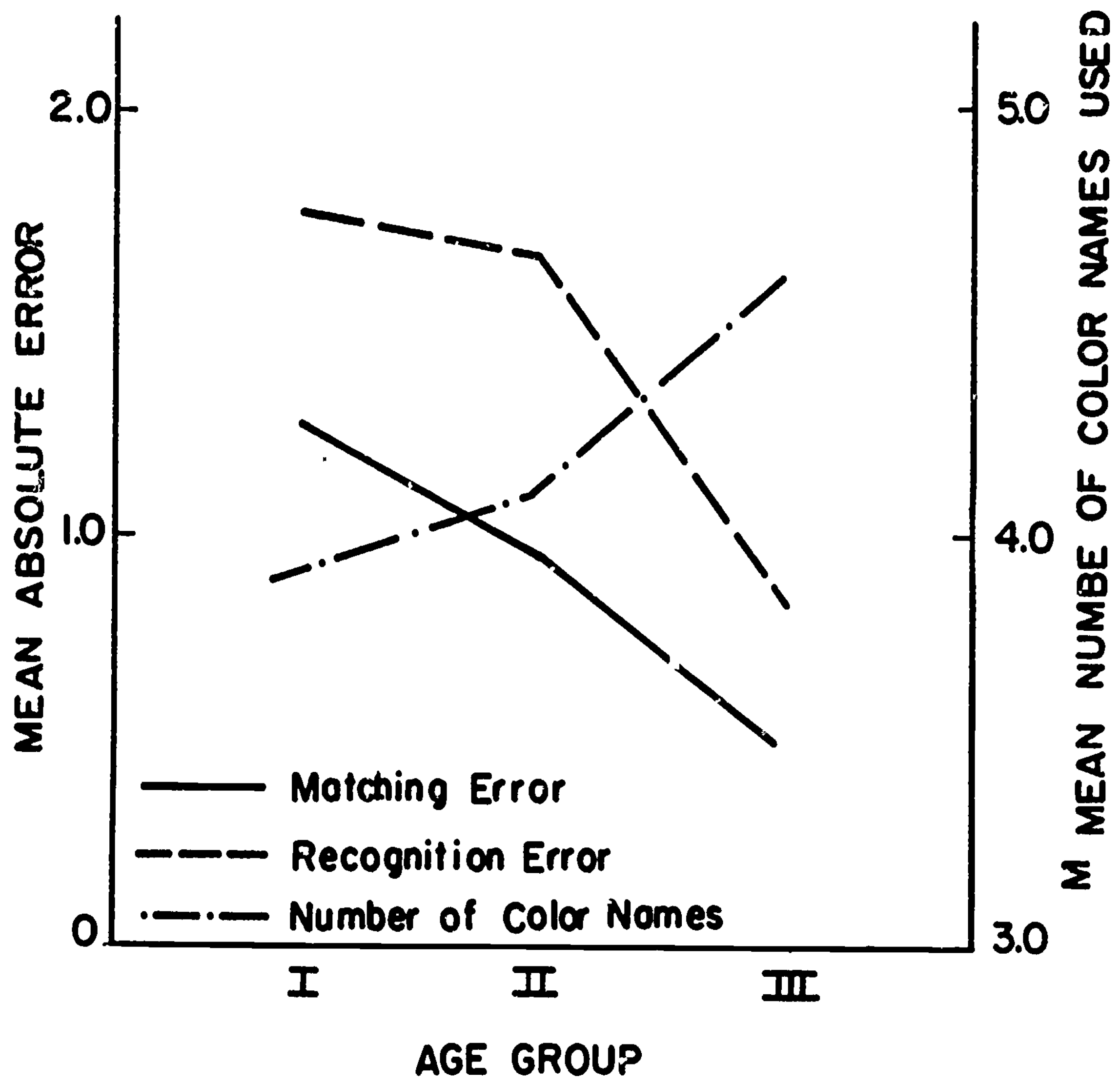

Figure 6. Development of Matching and Recognition Error and Color Names with Age.

APPENDIX A -- Stimulus Orders

Order I	Order II	Order III
3	10	2
8	1	11
14	7	3
10	3	9
4	11	1
9	2	5
2	9	10
6	13	6
1	8	13
11	4	8
5	12	14
13	6	4
7	14	12
12	5	7

APPENDIX B -- Data from Color Experiment

Subject	Age	Group	Task	Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	4;5	II	Match	I	1	2	3	3	9	7	7	7	9	11	12	12	13	13
			Name	II	PI	NN	NN	NN	NN	GR	GR	GR	BL	BL	BL	PU	PU	GR
			*Recog	III	1	5	1	10	14	10	7	12	10	4	10	8	13	9
			Fetch-BL									X	X	X	X			
			Fetch-GR				X		X		X	X						
			Fetch-PU													X		
2	4;4	II	Match	II	12	3	4	4	7	7	8	7	9	10	11	12	13	12
			Name	III	PI	BR	BR	GR	GR	GR	GR	GR	GR	BL	BL	NN	NN	NN
			Recog	I	1	3	4	3	8	7	7	7	10	9	10	14	13	13
			Fetch-GR								X	X	X					
			Fetch-BR		X	X	X											
			Fetch-BL										X	X	X			
3	4:4	II	Match	III	14	2	3	7	5	7	8	9	9	9	11	11	13	13
			Name	I	BR	BR	BR	GR	GR	GR	GR	BL	BL	BL	PU	PU	PU	PI
			Recog	II	14	2	4	5	5	5	10	13	9	9	10	12	12	13
			Fetch-GR					X	X	X	X							
			Fetch-BR		X	X	X											
			Fetch-PU												X	X	X	
4	4:4	II	Match	I	12	2	3	3	10	8	9	8	10	10	11	12	12	13
			Name	II	PI	PI	NN	NN	NN	NN	NN	NN	NN	NN	NN	PI	PI	PI
			Recog	III	13	5	11	9	3	9	14	8	3	6	4	10	3	1
			Fetch-PI		X											X	X	X
			Fetch-BL											X		X		
5	4;6	II	Match	II	12	3	5	6	7	8	8	8	9	10	8	13	12	12
			Name	III	BR	BR	BR	GR	GR	GR	GR	BL	BL	BL	BL	PI	PI	PI
			*Recog	I	13	2	2	6	6	6	5	9	9	10	9	10	1	13
			Fetch-GR						X	X	X	X						
			Fetch-BL										X	X	X			
			Fetch-PI		X												X	X
6	4;8	III	Match	III	1	2	3	4	5	7	8	8	8	9	12	12	13	13
			Name	I	NN	GY	NN	NN	NN	GR	GR	BL	BL	BL	PU	PU	PU	PU
			Recog	II	1	2	2	4	5	5	7	7	9	10	11	12	12	14
			Fetch-BL											X	X			
			Fetch-GR						X	X								
			Fetch-PU												X	X		
			Fetch-GY					X	X									
7	4;7	III	Match	I	1	3	3	5	5	8	6	7	10	10	11	12	13	14
			Name	I	RE	BR	BR	GR	GR	GR	GR	GR	GR	BL	PU	PU	PU	PU
			Recog	III	1	1	3	5	6	7	7	8	9	10	11	13	10	13
			Fetch-GR					X	X	X	X	X	X	X				
			Fetch-PU		X											X	X	X
			Fetch-BR			X	X											
8	4;4	I	Match	II	2	2	2	5	6	8	8	8	9	10	10	11	13	12
			Name	III	RE	BR	BK	GR	BK	BK	GR	BK	PI	RE	PI	BK	RE	BK
			Recog	I	2	3	2	3	2	7	2	10	9	11	10	12	12	10
			Fetch-RE		X	X	X	X	X	X	X		X	X				
			Fetch-GR					X	X	X								
			Fetch-BL			X	X											
9	4;11	III	Match	III	2	2	2	3	5	4	7	8	10	10	11	13	13	14
			Name	I	PI	NN	NN	GR	GR	GR	GR	BL	BL	BL	BL	PI	PI	PI
			Recog	II	2	2	2	4	5	4	7	9	9	10	10	13	14	14
			Fetch-BL										X	X	X			
			Fetch-GR						X	X	X							
			Fetch-PI		X												X	X
10	4;8	III	Match	I	1	3	4	4	3	7	7	7	10	10	10	10	14	14
			Name	II	PI	TA	TA	TA	TA	GR	GR	GR	BL	BL	BL	BL	PI	PI
			Recog	III	13	3	3	5	10	8	7	8	10	11	10	10	11	14
			Fetch-TA				X	X										
			Fetch-BL										X	X	X	X		
			Fetch-GR							X	X	X						
11	5;0	III	Match	II	14	1	2	6	9	7	6	9	9	11	12	11	13	13
			Name	III	GY	BR	BR	GR	GR	GR	GR	GR	BL	BL	BL	GY	GY	GY
			*Recog	I	1	3	3	4	7	7	8	5	10	10	11	10	14	13
			Fetch-GR						X	X	X							
			Fetch-BL										X	X	X		X	
			Fetch-GY		X												X	X
12	5;1	III	Match	III	14	2	2	4	5	6	7	8	9	10	11	12	13	14
			Name	I	PI	BR	BR	GY	GR	GR	GR	GR	GR	BL	GY	PU	PU	PI
			*Recog	II	14	2	3	4	7	7	9	8	3	10	11	11	11	4
			Fetch-GR						X	X	X	X	X					
			Fetch-PU										X	X	X	X		
			Fetch-GY			X	X									X		

APPENDIX B -- Continued

Subject	Age	Group	Task	Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14
13	4;11	III	Match	I	1	2	3	4	5	6	7	7	9	10	11	12	13	14
			Name	II	PI	BR	BR	GR	GR	GR	GR	GR	BL	BL	BL	PU	PU	PI
			Recog	III	3	2	1	3	5	8	8	7	9	10	11	2	10	14
			Fetch-BL										X	X	X	X		
			Fetch-BR		X	X	X											X
			Fetch-GR						X	X	X	X	X					
14	3;11	I	Match	II	1	2	3	3	7	9	8	9	9	9	11	12	13	13
			Name	I	NN	OR	OR	GR	GR	GR	GR	BL	BL	BL	PU	PU	PU	PU
			*Recog	III	1	1	3	9	12	9	9	8	8	11	11	11	11	13
			Fetch-PU								X	X	X	X	X	X	X	
			Fetch-GR								X	X	X	X				
			Fetch-BL							X	X	X	X					
15	4;8	III	Match	III	1	2	3	5	5	7	8	9	9	10	11	12	13	13
			Name	I	PI	SC	BR	GR	GR	GR	GR	BG	BG	BL	BL	PI	PI	PI
			*Recog	II	3	3	3	4	4	7	8	9	9	10	11	12	14	14
			Fetch-PI															X
			Fetch-GR					X	X	X	X							
			Fetch-BG									X	X					
16	4;4	I	Match	I	2	3	12	3	2	4	12	10	1	8	4	10	1	14
			Name	II	NN	BR	BR	BR	GR	GR	BL	BL	BL	BL	PU	PU	PU	PU
			Recog	III	2	2	3	4	5	12	2	14	9	11	11	12	13	3
			Fetch-BL									X	X	X	X			
			Fetch-GR						X	X	X							
			Fetch-PU												X	X	X	
17	4;2	I	Match	II	14	3	3	4	7	8	8	8	10	10	11	12	13	14
			Name	III	RE	BR	BR	GR	GR	GR	GR	BL	BL	BL	BL	PU	PU	RE
			Recog	I	14	3	3	4	7	8	8	9	8	9	11	13	14	13
			Fetch-GR							X	X							
			Fetch-BL										X	X	X			
			Fetch-RE														X	X
18	4;3	I	Match	III	1	3	3	5	4	5	7	7	9	10	10	12	13	14
			Name	I	BR	BR	BR	GR	GR	BL	PU	GR	GR	GR	BL	GR	PU	PU
			*Recog	II	1	3	3	10	7	5	6	5	10	9	10	13	12	13
			Fetch-GR						X	X	X	X	X	X				
			Fetch-PU		X	X											X	X
			Fetch-BR		X	X	X											X
19	4;2	I	Match	I	13	3	8	4	7	13	8	8	8	10	12	12	13	13
			Name	II	WH	WH	WH	WH	WH	WH	BL	BL	BL	BL	BL	BL	WH	WH
			*Recog	III	3	13	3	5	5	5	6	7	7	9	9	5	5	5
			Fetch-BL						X	X	X	X	X	X	X	X		
			Fetch-GR							X	X	X	X	X	X			
			Fetch-WH		X	X	X	X	X	X						X	X	X
20	4;2	I	Match	II	1	3	3	5	6	6	8	9	9	10	10	11	12	14
			Name	III	BR	BR	BR	GR	GR	GR	GR	BL	BL	BL	PU	PU	PU	PU
			*Recog	I	9	3	3	7	5	8	8	9	9	8	10	13	13	12
			Fetch-PU														X	X
			Fetch-GR						X	X								
			Fetch-BL								X	X	X	X				
21	4;3	I	Match	III	14	2	3	9	7	7	7	6	9	7	8	13	10	12
			Name	I	RE	RE	RE	NN	BL	BL	RE	BL	BL	RE	GR	BL	GR	GR
			Recog	II	13	2	3	10	13	5	13	1	14	10	12	7	13	1
			Fetch-BL										X	X	X		X	
			Fetch-RE							X			X	X	X	X		
			Fetch-GR							X	X	X	X					
22	4;4	II	Match	I	14	3	3	4	6	8	9	10	10	10	11	12	13	14
			Name	II	PI	BR	BR	GR	GR	GR	GR	BL	BL	BL	BL	PU	PU	PI
			Recog	III	1	3	3	10	3	9	9	9	9	10	10	12	13	13
			Fetch-BL										X	X				
			Fetch-GR						X	X	X	X	X					
			Fetch-BR				X											
23	4;4	II	Match	II	14	3	3	5	5	8	8	7	8	11	10	12	13	13
			Name	III	PI	BR	BR	GR	GR	GR	GR	GR	GR	BL	BL	PU	PU	PU
			*Recog	I	10	6	3	5	6	6	8	8	8	10	10	12	12	14
			Fetch-PU		X											X	X	X
			Fetch-GR						X	X	X	X	X					
			Fetch-BR			X	X											
24	4;6	II	Match	III	14	14	3	4	8	9	8	9	9	10	10	11	12	13
			Name	I	PU	BR	BR	GR	GR	GR	GR	BL	BL	BL	BL	PU	PU	PU
			*Recog	II	13	3	3	7	8	8	9	8	10	10	12	12	12	13
			Fetch-BL										X	X	X			
			Fetch-GR					X	X	X	X	X						
			Fetch-PU													X	X	X

*Spontaneous namer

BG Beige BL Blue GR Green NN No Name Given PI Pink RE Red TA Tan
BK Black BR Brown GY Gray OR Orange PU Purple SC Skin-Color WH White

Bibliography

Brown, R. W., and Lenneberg, E. H. A Study in Language and Cognition. J. Abnorm. Soc. Psychol., 1954, 49, 454-462.

Burnham, R. W., and Clark, J. R. A Test of Hue Memory. J. Appl. Psychol., 1955, 39, 164-172.

Carroll, J. B., and Casagrande, J. B. The Function of Language Classifications in Behavior. In Maccoby, E. E., Newcomb, T. M., and Hartley (Eds.), Readings in Social Psychology. (3rd ed.) New York: Holt, 1958, 18-31.

Farnsworth, D. The Farnsworth-Munsell 100 Hue and Dichotamous Test for Color Vision. J. Opt. Soc. Amer., 1943, 33, 568-576.

Fishman, J. A. A Systematization of the Whorfian Hypothesis. Behav. Sci., 1960, 5, 323-339.

Flavell, J. H., Beach, D. R., and Chinsky, J. M. Spontaneous Verbal Rehearsal in a Memory Task as a Function of Age. Child Dev., 1966 37, 283-299.

Hoijer, H. Cultural Implications of Some Navaho Linguistic Categories. In Hymes, D. (Ed.), Language in Culture and Society. New York: Harper and Row, 1964, 142-148.

Keeney, T. J., Canizzo, S. R., and Flavell, J. H. Spontaneous and Induced Verbal Rehearsal in a Recall Task. Unpublished Study, University of Minnesota, 1967.

Kopp, J. L. I Sees 'em as I calls 'em: Hue Discrimination and Hue Naming Across Cultures. Unpublished doctoral dissertation, University of Michigan, 1967.

Lane, H. L. A Behavioral Basis for the Polarity Principle in Linguistics. In Salzinger, K., and Salzinger, S. (Eds.), Research in Verbal Behavior and Some Neurophysiological Implications. New York: Academic Press, 1967, 79-96.

Lantz, D. and Lenneberg, E. H. Verbal Communication and Color Memory in the Deaf and Hearing. Child Dev., 1966, 37, 765-779.

Lantz, D. and Stefflre, V. Language and Cognition Revisited. J. Abnorm. Soc. Psychol., 1964, 69, 472-481.

Lenneberg, E. H. Cognition and Ethnolinguistics. Language, 1953, 29, 463-471.

Lenneberg, E. H. Color Naming, Color Recognition, Color Discrimination: a Re-appraisal. Percept. Motor Skills, 1961, 12, 375-382.

Lenneberg, E. H. Biological Foundations of Language. New York: John Wiley, 1967.

Lenneberg, E. H., and Roberts, J. M. The Language of Experience: A Study in Methodology. Memoir 13, Indiana University Publications in Anthropology and Linguistics, 1956.

Mathiot, M. Noun Classes and Folk Taxonomy in Papago. In Hymes, D. (Ed.), Language in Culture and Society. New York: Harper and Row, 1964, 154-161.

McNeill, D. Anthropological Psycholinguistics. Unpublished paper, University of Michigan, 1965.

Reese, H. W. Verbal Mediation as a Function of Age Level. Psychol. Bull., 1962, 15, 502-509.

Rivoire, J. L., and Kidd, A. H. The Development of Perception of Color, Space, and Movement in Children. In Kidd, A. H., and Rivoire, J. L. (Eds.), Perceptual Development in Children. New York: International Universities Press, Inc., 1966, 81-112.

OE-BR

Vygotsky, L. S. Thought and Language. Cambridge, Mass: The M.I.T. Press, 1962.

Whorf, B. L. Science and Linguistics. In Carroll, J. B. (Ed.), Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. Cambridge, Mass.: The M.I.T. Press, 1956, pp. 207-219.