Professional anthropologists also bring us reports of excellent memories, although very little systematic evidence relating to memory has been collected. For example, D. Reisman quotes the report of an anthropological colleague that among a remote people in the Philippines, "messages are conveyed orally . . . with an accuracy which is fabulous to us" (Reisman, 1956, p. 9). A similar point is made in a contrasting manner by Elizabeth Bowen in the example cited earlier in which she recounts the displeasure and consternation of her Nigerian hosts at her inability to remember the names of local plants which every ten year old in the village had long since committed to memory (Bowen, 1954, p. 16).

Other modes of inquiry support the anecdotal evidence, which suggests that members of a nonliterate, traditional society have developed mnemonic skills that are quite different from those of their literate, technologically advanced brethren. For example, philological and historical evidence led E. A. Havelock (1963) and others to maintain that an oral tradition produces special mnemonic devices, such as the epic poem, which function as an "oral encyclopedia" of the social, material, and historical aspects of the culture. This idea, recently popularized by Marshal McLuhan (1969), is echoed by Reisman when he suggests that members of a literate culture "can afford to be careless with the spoken word, backstopped as we are by the written one" (Reisman, 1956, p. 9). Nonliterates, unable to store their experience in print, must devote full attention to the spoken word. Reisman, in a manner similar to Havelock, adduced evidence that nonliterates (in this case New Guinea headhunters and Zuni shamans) have developed special mnemonic habits for the organization of cultural material.

The reports of anthropologists concerning the great importance that many tribal peoples place on learning of history, mythology, and traditions are consistent with Havelock's ideas about memory and literacy. For example, W. D'Azevedo reports that among the Gola of western Liberia, "An elder with a poor memory, or 'whose old people told him nothing' is a 'small boy' among the elders, and might well be looked upon with contempt by younger persons" (1962, p. 13).

In addition to philological and anthropological evidence, there is a very small amount of experimental, or quasi-experimental, evidence on the question of culture and memory gathered by psychologists. One such source is the evidence reported by the IQ testers, who frequently note that subtests depending upon oral or visual memory produce scores that are equivalent to, or in some cases superior to, Western norms,

while subtests in which memory is not an important factor suffer by comparison. Analogous findings have been reported by L. Doob (1965) in a study of the ability to recall visual stimuli. His data indicate that eidetic imagery, or the ability to recall visual stimuli exactly, is encountered far more frequently in Africa than in the United States.

One quasi-experimental psychological investigation of nonliterate peoples was carried out by F. C. Bartlett (1932) among Swazi of South Africa. Having heard of the "marvelous word-perfect memory of the Swazi from his childhood up" (p. 248), Bartlett set out to find out under what conditions this phenomenal memory manifested itself. First he asked a young boy to carry a message to someone else in the village, and found that recall was comparable to that which more systematic experiments had shown for English children of similar age. He then tested a cowherder's memory for a series of transactions involving cattle that had been sold the year before. In this case, the herder's memory was found to be phenomenally accurate, although he had been only peripherally involved in the transaction. Bartlett attributed the herder's performance to the importance of cattle as a medium of exchange among the Swazi, and suggested that because of this "persistent social tendency," the performance was really not so remarkable. The cowherder's feat of memory seemed outstanding because what was socially important to him was irrelevant to the Western observer, who therefore found a good memory for cows and prices quite unusual. In fact, we might expect the Swazi cowherder to be equally astounded should he encounter two American ten year olds trading baseball cards with the intricate recall of players, teams, batting averages, and relative standings that a successful trader requires.

The many hypotheses that can be generated from Bartlett's demonstration have never been systematically followed up and tested. For example, in what specific ways does a "persistent social tendency" influence recall? Does it produce different *ways* of recalling as well as different *amounts* recalled? Bartlett himself, when comparing the Swazi to the Westerner, suggested that culture determines a difference in the way things are recalled. He hypothesized that rote memory is the preferred memory technique of nonliterate people and defined rote memory as serial memorizing. He concluded:

> According to the general theory of remembering which has been put forward, there is a low level type of recall which comes as nearly as possible to what is often called rote recapitulation. It is characteristic of a mental life having relatively few interests, all somewhat concrete in character and no

one of which is dominant. Is there anything in social organization which parallels this state of affairs in mental organization and so, on the social side, favors the rote recapitulatory method? I think there is, and it is largely to this that we must look for the explanation of the reputation for excessively accurate and detailed memory which the more or less primitive group possesses. [Bartlett, 1932, p. 264]

Unfortunately, Bartlett's research has had little impact on subsequent research. S. F. Nadel (1937) provided evidence that themes of great cultural interest are best remembered, but G. Bateson (1958) provided anthropological evidence that serial recall is not characteristic of primitive people in general.

In our opinion, Bartlett's phrase "persistent social tendencies" has at least three interpretations, which have not been sorted out either conceptually or experimentally: (1) there are different levels of interest and motivation; (2) there are particular memory skills that different environmental conditions might produce; and (3) there are differences in the extent and ease of use of relevant vocabulary. Any one, or a combination, of these factors could account for Bartlett's results.

As described in Chapter 2, we too have noted a heavy reliance on what appeared to be serial rote learning in the classroom (see also Gay and Cole, 1967, pp. 33ff.). Students often copied exactly and step-by-step what the teacher said or wrote and failed completely to grasp the principle involved.

Our few observations among the Kpelle, combined with the anthropologists' casual observations and the psychologists' few experiments, led us to attempt a detailed and systematic experimental investigation in the hope of isolating those factors that influence the ways in which members of various cultures use memory as a cognitive tool. In order to move beyond the level of casual observations, we had first to choose an experimental tool or set of tools that would more nearly fit our idea of a memory problem than the concept-discrimination studies. The experimental technique should be flexible enough to make possible the study of memory in nonliterate societies, and at the same time should enable us to evaluate hypotheses concerning the particular memory skills being used by our subjects.

The task we selected for detailed experimentation, the free-recall experiment, has several features that render it useful for our purposes. First, it is extremely easy to administer. A subject is presented a series of items, one at a time, and is told that he must try to learn them so that he can recall them at a later time. After the last item is presented,

a period is given for recall. The list can then be repeated as many times as the experimenter wishes. Second, the task is unstructured; the subject is free to remember in any manner he chooses, and the order in which subjects recall items gives important insight into the mechanisms of memory. W. A. Bousfield and his associates (Bousfield, 1953; Cohen, 1963) stimulated interest in this procedure by demonstrating that when the items to be remembered came from easily identifiable semantic categories, recall tended to be "clustered" so that items from a given semantic category were commonly recalled together. More recently, E. Tulving (1966) has measured the organization of recall in terms of the consistency between successive attempts by one subject to recall the same list. Although many questions of fact and theory remain to be clarified, it is clear from the work of these and other investigators that North American high-school and college students show a strong tendency to reorganize material presented for memorization and that success in recall is related to the degree of organization the subject imposes on the to-be-recalled list (see summary article by Tulving, 1968).

Although it might seem a contradiction at first glance to employ a memory task to study cognition, the concern with the organizational features of free recall fits nicely with definitions that emphasize the constructive and organizational features of cognition. In terms of our discussion in Chapter 1 (pp. 19–20), our interest in free recall could be characterized as a concern with the extent to which cognition plays a role in the memory process of different cultural groups.

Procedures

The basic procedure in each of the free-recall experiments (all essential modifications of procedure will be discussed as they occur) was for the experimenter to read the list of items to be recalled at a rate of approximately two seconds per item. After the entire list had been presented, the subject was asked to repeat as many of the items as he could. Approximately two minutes were permitted for recall, during which time the experimenter recorded each word on a specially prepared data sheet. The subject usually indicated prior to the end of the two minutes that he could recall no more and the next trial was then begun.

Unless otherwise specified, each subject was presented the same list five times but in a different order each time. The only restriction was that no two items from the same category occur adjacent to each other within a trial. Furthermore, the order in which different list orders were used differed from subject to subject, thereby randomizing the effects of list order.

For the basic series of experiments the list of items to be remembered was composed of the twenty clusterable items contained in Table 3–2. When desired for comparative purposes, the nonclusterable list from Table 3–2 was used.

Measures of Performance

In justifying the selection of free-recall task as a method for studying memory, we emphasized the great freedom permitted the memorizer in structuring his recall. It is understandable, then, that our measures of memory will represent alternative ways of assessing different kinds of structure.

An example of one possible structure is implied by Bartlett's hypothesis that rote recapitulation characterizes the recall of traditional, nonliterate peoples. As he describes the process, rote memory entails always beginning recall from the beginning of the to-be-recalled sequence. An analogue of this theory for the free-recall task would be the case where the subject remembers the items in the same order that they were presented by the experimenter.

In order to measure this serializing tendency, we calculated the correlation between the order of the experimenter's presentation of the to-be-remembered items and the order of the subject's recall. This correlation statistic (Pearson's r) then became a datum characterizing the degree of one kind of recall structure for a given trial. It was possible to compare different groups on the amount of serial organization under various conditions.

A major alternative to serial organization is clustering, the tendency to group items that are part of the same class together in the recall list. The nature of the particular class can be defined in various ways, and recall lists can be evaluated for the degree to which the observed clustering exceeds the amount expected if the items had been drawn at random from the to-be-remembered list (semantic classes are the focus of most of our attention, but functional classes, for instance, could be studied).

The measure of clustering used in our work is a "standard deviate" (z score), which is a measure of the extent to which a particular recall list deviates from chance clustering. A perfectly random list corresponds to

a z score of zero. Clustering is reflected in positive z scores. Negative z scores are also possible and reflect systematic organization that runs counter to clustering (such would be the case if the subject is showing perfect serial organization). A discussion of the measurement of organization in free recall is contained in Frankel and Cole (1971).

Finally, we consider the question of the amount of recall. The simplest measure of how much is remembered is simply the number, or proportion, of items recalled. We, of course, present this basic datum and, as we shall see, the relation between practice with a particular recall list and the number recalled will be a central problem for analysis.

In addition to total number recalled, we will also include analysis of the number recalled from different parts of the to-be-remembered list. In free-recall studies in the United States (Deese, 1957; Cole, Frankel, and Sharp, 1971) one typically observes a "serial position effect"; items near the beginning or end of to-be-remembered list are better recalled than those in the middle. This fact is widely interpreted (cf. Atkinson and Shiffrin, 1968) as evidence for the presence of two distinct memory processes: a short-term process (reflected in near-perfect recall of words from the end of the list) and a long-term process (reflected in superior recall of words from the beginning of the list). Since group differences may be localized in a particular part of the list (for example Cole, Frankel, and Sharp, 1971, found that older schoolchildren remembered more items than younger schoolchildren only in the early and middle positions of the list), serial-position analysis offers still another measure of structuring in memory processes. The inference of short-term and long-term processes from the kinds of structure that are involved in this measure of recall is still a very controversial matter, but the universality of the fact that differential recall is observed for different serial positions in American studies suggests the usefulness of including such analyses for cross-cultural comparisons.

Keeping in mind the fact that our measures of performance represent various indicators of possible cognitive processes entering into memory performance, we turn to the first of our studies.

Experiment 1: Are Clusterable Lists Easier to Learn Than Nonclusterable Ones?

The opinion is fairly widespread among American psychologists who study memory that recall and organization are closely related; the better the to-be-remembered material is organized, the better it will be recalled (Mandler, 1966; Tulving, 1968). A prediction that follows from this generalization is that all other things being equal, clusterable lists should be easier to recall than nonclusterable, randomly constituted lists. C. Cofer reviewed the meager evidence up until 1966 and concluded that clusterable lists are in general easier to learn than nonclusterable lists (Cofer, 1967, pp. 181ff.). The cross-cultural generality of this finding was the subject of our first recall experiment.

SUBJECTS AND PROCEDURES

The subjects in this experiment all lived in the area of Cuttington College. Twenty subjects were obtained in each of the basic population groups: nonliterate six- to eight-year-olds, nonliterate ten- to fourteen-year-olds (first grade), and school ten- to fourteen-year-olds (second to fourth grade). Half of the subjects in each group were presented a clusterable list and half a nonclusterable list (Table 3–2). Using the standard procedure outlined above, each subject was given five recall trials of an orally presented list.

The experimental design thus includes comparisons of age, education, and "clusterability" of the stimulus list. Details of the procedure for this and the other standard free-recall experiments are given in Appendix E.

RESULTS

Each factor of concern in this first experiment affected the number of items recalled. For all subject populations, the clusterable list was slightly easier to recall than the nonclusterable list (8.7 items per trial versus 7.6 items). Recall increases slightly as a function of both age and education as shown in Figure 4–4. These results are shown as an average across trials because group differences were approximately the same at all stages of learning.

Two conspicuous aspects of the groups' recall performance, which are not represented in Figure 4–4, require comment. First of all, an average of 7.0 items was recalled on Trial 1 and recall increased only to 8.8 on Trial 5. Although this increase is statistically reliable, its magnitude is smaller than one would anticipate on the basis of previously reported American data. Second, there is very little variation in accuracy among items as a function of their position in the recall list. We reported superior performance near the beginning of the list in our paired-associate study. But no strong effects of this kind are evident among our Kpelle subjects given the free-recall task. Figure 4–5 shows

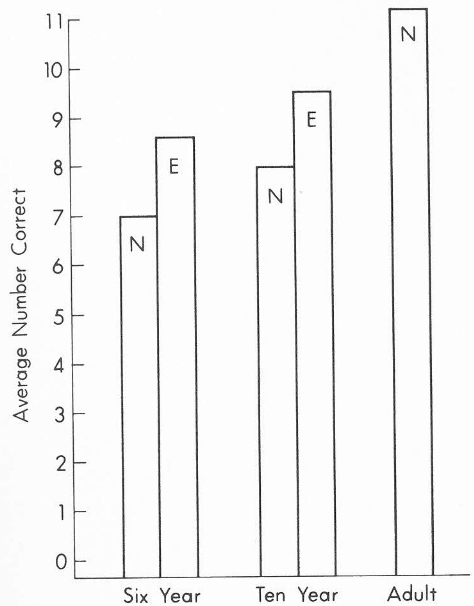

FIGURE 4–4 The Average Number of Items Recalled per Trial as a Function of Age and Education (E = educated, N = noneducated)

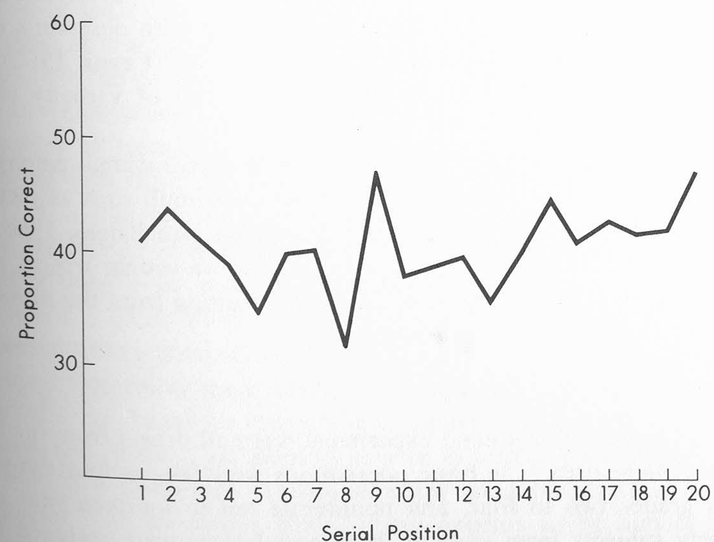

FIGURE 4–5 Relation between Accuracy and Serial Position of an Item during Presentation

the relation between accuracy and the serial position of the items at the time of presentation to be relatively flat.

The Kpelle also failed to organize recall according to semantic categories in the clusterable conditions where such clustering was possible. The average clustering z score was $-.13$ for all five groups taken as a whole; there was no significant variation among the groups. A similar lack of organization is reflected in the seriation measure. Far from showing rote learning, none of the groups studied showed any significant correlation between presentation and recall sequences, and the average for the experiment was $r = -.05$.

Thus, it would appear that for the populations studied, initial recall was anything but impressive, increases in recall with practice were negligible, and conspicuous organization either in terms of the presentation sequence or semantic properties of the list was absent. Moreover, the difference among populations and conditions, while reliable, were not of great magnitude.

Experiment 2: Do the Type of Stimulus Materials and Manner in Which the Presentation Lists Are Organized Affect Recall?

One of the first hypotheses that the relatively poor performance in Experiment 1 suggested (an hypothesis consistent with observer's comments about the "concrete mentality of the African," Cryns, 1962) was that presentation of the stimuli concretely instead of verbally would greatly enhance recall.

Evaluation of the relative effectiveness of concrete stimuli among the Kpelle is complicated by the fact that concrete stimuli such as pictures or objects are more easily recalled by Americans (see Paivio, 1968, for a summary of these data). Consequently, what we sought was to measure the *relative* amounts of improvements resulting from the introduction of concrete stimuli in the two cultures.

SUBJECTS AND PROCEDURES

The subjects in this second experiment were all drawn from the Cuttington College area. The basic populations were ten- to fourteen-year-olds in grades two to four, and nonliterate ten to fourteen year olds. The forty subjects from each of these populations were assigned haphazardly to four different experimental treatments, representing all combinations of two kinds of stimulus materials (spoken words or objects) and two ways of ordering the presentation lists. For all groups the lists consisted of the twenty items from Table 3–2 used in the clusterable groups of Experiment 1. For half of the subjects, this meant that no item from a given semantic class ever occurred next to another item from that class. For the remaining half of the subjects, the presentation orders were "blocked"; that is, items from within a semantic class always occurred together in the list. Blocks of items were arranged differently on each trial in a random fashion.

To summarize, the basic comparisons included in this experiment were educated versus nonliterate subjects, random versus blocked presentation orders, and objects versus words as stimuli.

RESULTS AND DISCUSSION

As was true in Experiment 1, our manipulations of the conditions for remembering produced only small effects on the average number of items recalled. There was no reliable difference between the children who had attended school and those who had not; there was only a slight advantage for the object over the word stimuli and for blocked over random presentation. Consistent with Experiment 1 results, we find no tendency for subjects to recall in serial order at any time in the training ($r = -.07$). There were, however, indications that the organization of recall differed between the educated and nonliterate groups and as a function of the experimental conditions (see Appendix E for a detailed discussion of these results). First, the educated subjects manifested a serial-position effect similar to that observed with American children. Second, we observed a significant amount of semantic clustering during the course of learning for certain of the groups. The development of clustering with successive trials is shown in Figure 4–6. It is clear from Figure 4–6 that for the blocked conditions, clustering begins at well above the chance level, while for the random groups, clustering only approaches nonchance levels toward the end of training. Presenting objects has a marked effect on clustering only in the case where stimuli are presented in a blocked order.

Several features of these results seem unusual if one uses typical data collected from American subjects as a reference point. The items seem common and distinct, yet memory is poor in terms of numbers recalled. Not only is performance initially at a low level, but there is little improvement with successive trials. There appears to be almost a total lack of semantic clustering except under very favorable circumstances (blocked presentation of objects) and virtually no relation between se-

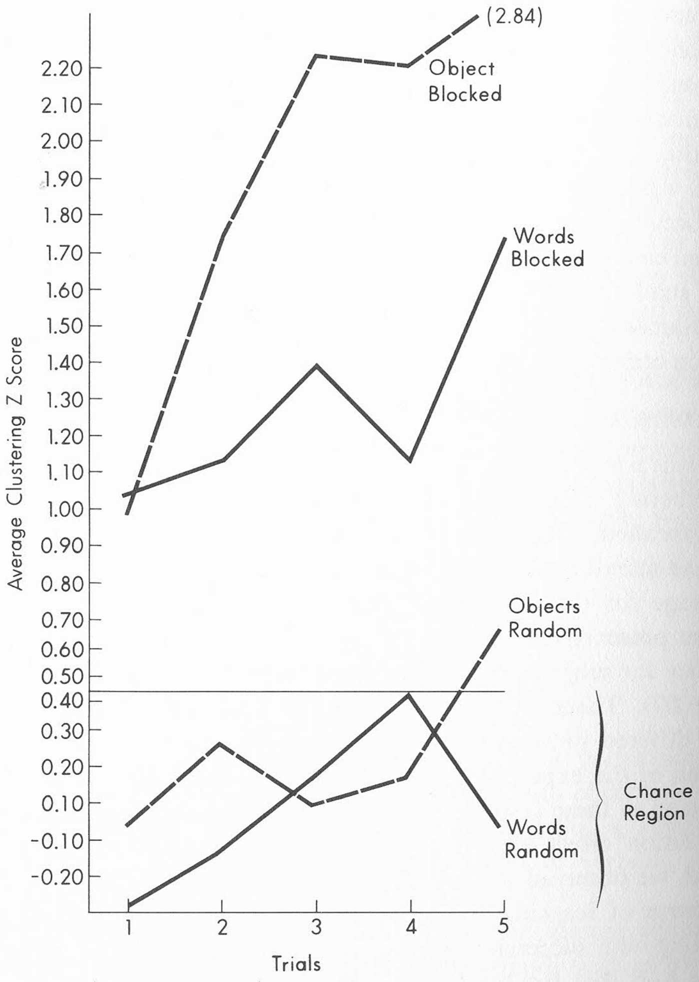

FIGURE 4–6 Average Clustering Score for Each Trial as a Function of Presentation Orders and Stimuli Presented. (The solid line at .44 represents the average score which reliably exceeds chance.)

mantic organization and amount recalled. Furthermore, alternative measures of organization (seriation and differential recall of items from various parts of the list) generally failed to indicate structure in recall.

In order to untangle the many factors that could be controlling this pattern of performance, we conducted a large series of studies, each of which was designed to evaluate a different hypothesis about organization and memory.

As one starting point in this research program, we conducted studies with various groups of American schoolchildren and young adults to obtain a clearer picture of the features of free recall that result from experimental manipulations such as those applied in Experiments 1 and 2 among the Kpelle. Because of several procedural variations that were introduced in Kpelleland, we undertook the studies in the United States to determine the influence of such variables as list clusterability, blocked-presentation order, and words versus objects as stimuli. The stimuli used in these studies were taken from Table 3–5, containing the lists used in our study of free-association responses. In addition, we collected data from groups of Mexican Indians living in Yucatan and another tribal group in Liberia. These latter data are included in Appendix E.

American Free Recall

The main features of recall performance among our American subjects are the following (see Cole, Frankel, and Sharp, 1971, for a detailed exposition):

1. For children in grades one to eight the number of items recalled on the first recall trial is comparable to our Kpelle results (seven to ten of the items). However, except for the youngest children, there is more improvement over trials than observed in Experiments 1 or 2.
2. College students recall many more words, and recall is essentially perfect by Trial 2.
3. All American groups are more sensitive to the serial position of the item to be recalled; averaging over trials, the last item presented is correct about 80 percent of the time across all groups. Recall scores for the different groups differ only in other positions. Trial 1 differs from the remaining trials in that on Trial 1 there is a large primacy effect (items near the beginning of the list are best remembered), but there is little recency (recall of late items from the list). On later trials, recency dominates.
4. On Trial 1 there is a significant positive correlation between presentation and recall order. On the remaining trials, this correlation is negative. The Americans begin by trying to rote memorize the list!
5. The relative increase in recall as a result of presenting objects instead of words is even greater than we observed among the Kpelle.
6. Except for the youngest children, there is considerably more semantic clustering among the American students. College students show almost perfect clustering in a very few trials.

In summary, there is an orderly development of free-recall learning among our American subjects. By the third grade, average performance is roughly at the level observed by our various African groups, but thereafter, performance among the Americans shows better recall and organization.

Summary of the Preliminary Experiments

The picture that emerges from these studies of free-recall performance in Liberia, the United States and Mexico consists of a set of regularities with some important divergences.

In all of the cultures sampled, the variables that control free recall seem to operate in very similar ways. There was a slight tendency for clusterable lists to produce better recall than nonclusterable lists, but this difference was statistically reliable only in the case of the Kpelle. Presentation of objects instead of words enhanced recall in each culture where it was tried. The same was true for the presentation of lists in a blocked rather than a random order.

A significant difference in the patterns of performance occurs when we compare changes in the Kpelle performance from Trial 1 to Trial 5 with changes observed for our older American subjects. The absolute level of recall would be very similar across cultures if we only considered Trial 1. Where large cultural differences in amount recalled are observed, the largest differences occur following Trial 1 and increase across trials.

In contrast with the number of items remembered, the way in which recall is *organized* differs across cultures from the outset of training. In general, the American subjects show primacy on Trial 1 (accompanied by a positive correlation between presentation and recall orders) and recency thereafter. There is also a significant amount of clustering from the outset, except for the youngest American groups. High levels of clustering and trial-related changes in serial-position responding are generally absent from the Kpelle data. For these groups conditions that are relatively favorable to recall and organization have their effect on Trial 1.

Hence, we return to a consideration of memory among the Kpelle with an orientation that differs significantly from the hypotheses that we started with. Considerations of "concrete versus abstract" learning, and ignorance about the major features of free recall among nonliterate groups have given way to a more precisely defined inquiry. Now we seek to determine why there is relatively little recall or clustering among Kpelle subjects, even after repeated practice. Are there experimental or

naturally occurring occasions upon which the Kpelle will exhibit recall of the same quality as that typically observed in the United States?

Our approach to answering these and related questions can be divided roughly into two categories. First, we instituted a series of studies that were similar to Experiment 1 in terms of the basic procedures, but that varied characteristics of the subjects or the general conditions of the experiment. These studies were directed at such hypotheses as: perhaps the task seemed unimportant to the Kpelle subjects, so they were not trying; or perhaps one needs to have several years of schooling before free recall becomes organized.

Following our evaluation of hypotheses of this type, which involve no fundamental changes in procedure, we turn to a series of studies aimed at changing the basic *structure* of the free-recall tasks. The intent of these studies is to find ways of presenting the task that will evoke efficient, organized performance. After an analysis of the way performance depends on the structure of the task, we will try to make some educated guesses about the cultural factors influencing free-recall memory.

Experiment 3: The Effects of Different Motivating Conditions on Recall among the Kpelle

As an example of how one might come to consider lack of interest on the part of our Kpelle subjects as a determinant of their performance, consider the phenomenon of a *kwii* Liberian college student wandering into a hinterland village. This event is a little unusual, but hardly a matter of great moment to the villagers. If the student is a local boy, he is apt to be met somewhat patronizingly. The village adults will respect his book learning but will still consider him a "small boy" in the important matters of life. If he is a stranger, he might be met with some suspicion; and if he is thought to be a tax collector or government agent, suspicion might easily turn into enmity.

When the college student explains that he is visiting the town in order to talk with the people, it is likely to be some time before he can convince them that the outcome of this talk is likely to be harmless. It is emphasized that the project directors are teachers, interested in helping the children "learn book." Whenever possible, the town chief and his council of elders are consulted and shown traditional courtesies.

A possible outcome of this emphasis on the lack of a connection between the experimenter and the government, as well as the generally harmless nature of the tasks involved, might be to prevent the subjects from taking the experiment seriously. Although subjects were "dashed" a can of fish or a small amount of money for their cooperation, the knowledge that the outcome was relevant only to some far-off schoolteacher could not have been of great concern. These doubts seem all the more plausible when we contrast the view of this would-be Kpelle subject with that of an American schoolchild, whose response to the experimental situation is likely to be, "Is this an intelligence test?" and whose desire to exhibit his intelligence often produces overt signs of anxiety.

Consequently, it was decided to determine the effect of providing monetary incentives for good performance on recall under certain of our standard conditions. Two such experiments were conducted at about the same time.

In the first incentive experiment, half of the subjects were told that they would receive at least thirty-five cents for their participation, but that they could earn up to twenty-five cents more if they performed well. The remainder were paid a flat thirty-five cents. Since rural Kpelle consider seventy-five cents a good wage for a full day's work and many workers receive only fifty cents a day, the promise of up to sixty cents for twenty minutes of a man's time was thought to be an adequate incentive.

The four groups in this experiment, each consisting of ten nonliterate adult subjects, represented the factorial combination of two incentive conditions (incentive versus no incentive) and two kinds of lists (nonclusterable and clusterable). In all other respects, the procedures were exactly like those used in the standard experiments.

The results conformed to those obtained earlier and there was no difference between the incentive and no-incentive conditions. If the Kpelle performances in the previous experiments were the result of motivational deficiencies, the incentive motivation used in this experiment was clearly inadequate.

A second experiment explored a slightly different motivating manipulation. Instead of simply telling the subject that he could earn more money by recalling more items, the subject was given a running account of his performance by the use of pebbles, which represented money. Each time that a word was correctly recalled, a pebble was added to the subject's pile, and he was told that he would get a penny for every four pebbles (five pebbles in the case of children). Pebbles are traditional markers used as counting devices in many situations and hence their function was thought to be clear in the context.

The results were, in all essential respects, the same as those obtained

in the original free-recall experiment. There were small differences in the numbers of items recalled by the various groups, with the younger children performing worst. There was no noticeable improvement across trials and there was no reliable difference among the older children and the adults. Overall, there was no significant clustering.

One indication that subjects were not unaware of the consequences of increasing the number of items recalled was a marked tendency on the part of several subjects to say a great many items in their recall lists. The number of intrusions from repeated items or responses not on the original list was greater than we observed on other occasions.

On balance we can conclude that we have once again failed to affect the course of learning through a change in the incentives offered, and it seems that no qualitative changes are to be expected from this source. Granting the possibility that some other motivating manipulation might prove effective, we moved on to assess other plausible variables that might be at work.

Experiment 4: Recall among Nontraditional Kpelle Groups

One question that quickly comes to mind is whether or not Kpelle whose life experiences have taken them far outside traditional Kpelle culture will manifest the same kinds of recall phenomena as those described thus far. The data from the Vai (Appendix E) suggest that literacy or degree of Westernization might affect memory performance. The data from the second- through sixth-grade students seem to contradict such a conclusion, but further exploration is clearly in order. Initial steps toward answering this question were undertaken in three small studies.

Our first experiment on the recall of nontraditional Kpelle subjects compared ten- to fourteen-year-old schoolchildren who lived either in the area of Cuttington College or in Monrovia. We used schoolchildren in both groups so that only the fact of living in an urban setting or a semitraditional village would distinguish the two groups. The procedure used was that of the clusterable groups in the first free-recall experiment. The results are easily summarized. The urban Kpelle remembered slightly, but reliably more than the rural Kpelle (10.2 versus 8.7 items per trial, respectively). In all other respects, the data were virtually identical to the results of Experiment 1. There was little improvement over trials and no significant clustering. The magnitude of this difference is about equal to that between the Vai who knew Vai script and those who did not, suggesting that something to do with modernization is leading to the increase in recall. However, the lack of clustering among the Kpelle cautions us against judging the similarity of the underlying mechanisms, and in any event the major features of the data are quite unlike those observed in our studies in the United States.

In the next study, the clusterable list was presented to two groups of eighteen- to twenty-year-old Kpelle. One group consisted of ten nonliterate Kpelle from the town of Salayea, approximately sixty miles northwest of Cuttington College. Salayea is located on the main all-weather road running from Monrovia to the Sierra Leone border, but like the residents of Sinyee, these people were relatively traditional Kpelle rice farmers. The remaining ten subjects were Kpelle students attending two nearby high schools, the Lutheran Training Institute and Zorzor Training Institute. The students were in grades ten through twelve; in general they were living away from home, and as indicated in Chapter 2, they represent a nontypical population.

A second study contrasting nonliterate and high-school-educated subjects was conducted in the Cuttington area. An additional feature of this experiment was that training was continued for fifteen trials, instead of the usual five trials, in order to determine the effect of really extensive practice on recall and organization. The subjects were eighteen to twenty years of age. The high-school students were attending school in the county administrative center, Gbarnga. The nonliterate subjects lived in the town of Galai, a "feeder" town, many of whose citizens worked at Cuttington College. The first of these two studies was conducted by John Kellemu, the second by Paul Ricks. Except for the number of training trials given, the procedures for these two studies were identical to those used for the clusterable groups given oral presentation of the lists in the previous experiments.

The results of both experiments in terms of number recalled per trial are shown in Figure 4–7. The two experiments were consistent in showing a sizable superiority of the high-school students over the nonliterate subjects. This difference, significant from the outset, increased over trials; the high schoolers continued to improve with training, but the nonliterate subjects showed no improvement after Trial 3, and little improvement overall.

The same general relation between the two populations was observed in their clustering scores. As shown in Figure 4–8, clustering was similar for the two groups early in training, but the high schoolers show a

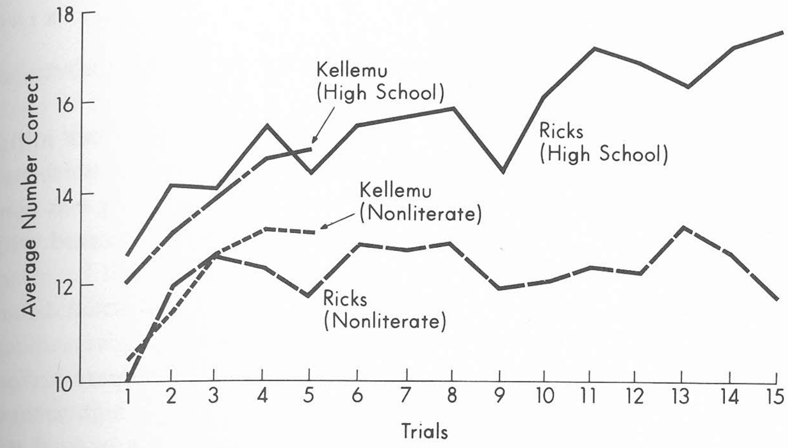

FIGURE 4–7 Average Number of Items Recalled per Trial for High School and Non-Educated Subjects. Replications by Ricks and Kellemu

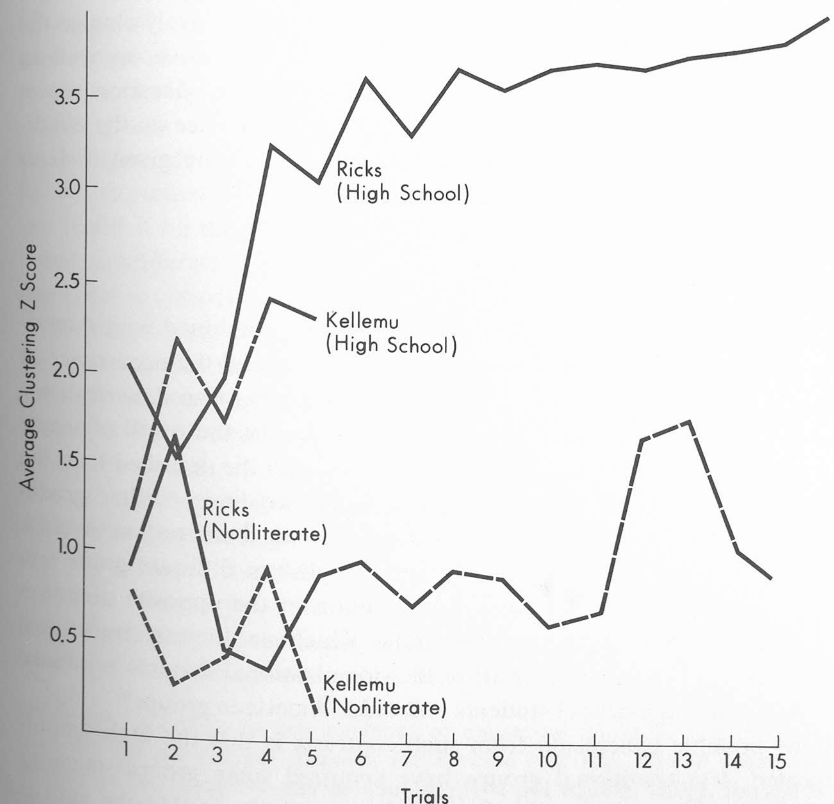

FIGURE 4–8 Average Clustering Scores as a Function of Trials Corresponding to Accuracy Data in Figure 4–7

clear improvement over trials, whereas there is little or no improvement for the nonliterate subjects for many trials.

In one general respect, these data differ from the data in our initial experiments; even for the nonliterate subjects, recall was a little better (twelve instead of nine items recalled per trial), and clustering was considerably better (.56 instead of −.13) than previously observed with nonliterate subjects run under the same conditions.

These discrepancies remind us once again that relatively small difference in the absolute levels of performance should not be given undue weight. It also reminds us that experiments that seek to compare performance for groups having dissimilar past experiences (like high-school-educated subjects versus nonliterates) must include both groups if the effect is not to be attributable to the experimenter or some other unknown factor, rather than to the variable in which one is interested.

On the basis of this scanty evidence, we can tentatively conclude that simply living in an urban environment does not qualitatively change the major features of free-recall learning, but that qualitative, as well as quantitative, changes in learning occur at the level of education represented by our high-school subjects. Some further evidence on the conditions under which education influences memory will be given in later experiments in this series.

What Produces Good Recall?

The results with high-school-educated Kpelle, combined with American data showing that increasing recall over trials and the occurrence of semantic organization begin after about three years of an American education, suggest that some features of recall may be the result of something connected with literacy when combined with the detached learning of the book and schoolroom. Unfortunately, we have neither grade-by-grade data for the Liberian student nor an independent evaluation of their degree of fluency in written English at different grade levels. Consequently, we will turn our attention in the opposite direction and ask: are there circumstances under which nonliterate, traditional people will manifest some or all of the organizational features produced by the Kpelle high-school students and older American groups?

One possible conclusion from such contrasts is that the high-school-educated, less traditional groups have acquired some general memory skill which the nonliterates lack. Another possibility is that the skill is in some sense specific; the nonliterate will manifest the same sorts of recall as the high-school student but under different conditions. It is these kinds of hypotheses that we sought to evaluate through a set of experiments that manipulate various facets of the recall situation in order to determine their effect on the recall of traditional subjects.

In two of the experiments attacking this problem we stumbled upon these two facts: (1) increasing the number of items to be remembered increased the number recalled; (2) asking the subject to place the items in a bucket, or to sort the items into cups, had a large effect on the number recalled and greatly increased the amount of semantic clustering. In the latter study, we seemed to have hit upon a mechanism for making the recall performance of the nonliterate Kpelle approximate the kind of recall we have observed in literate groups. The most likely candidate for the cause of the improved recall was the fact that subjects manipulated the items that were said to "go with the cups." Perhaps the cups served in some way to remind the subject of the items. Put in the language of contemporary theorizing about memory, these speculations give rise to the hypothesis that the cups act as cues, which aid the subject in retrieving the items from his memory.

The question then occurs: can the differences in amount and organization of recall that we have encountered in the studies reported thus far be accounted for on the basis of different retrieval cues? In particular, could it be the case that given the proper retrieval cues, the traditional, nonliterate Kpelle will show the organizational and recall features that we associate with our American subjects?

G. Mandler raises much the same issues in his recent discussions of organization and memory (Mandler, 1966). For example, he points out that it is necessary to make a distinction between the use of rules for effective retrieval from memory (such as semantic categories) and their discovery. In fact, Mandler goes so far as to hypothesize that the function of repeated trials in free-recall experiments, such as those we have been describing, is to give the subject repeated opportunities to discover the rule latent in the lists. Thus he says, "The free-recall situation demands a discovery by the subjects of some adequate rules that will allow them to make the input items accessible, and to retrieve them adequately" (Mandler, 1966, p. 41). Of clustering he says:

> Obviously, a failure to find evidence of clustering might be due to the fact that the subject did not discover the specific rule that related members of subsets of the input list one or another. It is not equally likely, though it is possible, that subjects might discover the rule without being able to use it adequately. [Mandler, 1966, pp. 40–41]

It would seem that both in terms of an explanation of the results we observed with our Kpelle subjects and on more general theoretical grounds, it behooves us to make a systematic investigation of what kinds of cues might facilitate the use of, or the discovery and use of, retrieval rules.

Experiment 5: The Identification of To-Be-Remembered Items with External Objects: The Chairs Experiment

One of the first experiments we designed on the question of cues for recall was invented in the context of a discussion of the term *concrete.* It will be recalled that in introducing the contrast between objects and words as stimuli, we noted the suggestion that Africans have "concrete mentalities," and from this, we derived the idea that the use of concrete objects ought to enhance recall. When we failed to obtain the expected amount of improvement from the introduction of objects, one of our responses was to review the vague notion of a "concrete" stimulus and to hypothesize that what might be important is not the concreteness of the item to be learned, per se, but rather that the item have a concrete tie with some external object. In Mandler's terms, the subject might require such a concrete connection in order to discover the rule latent in the material. This idea fits in with our observations about sorting items into cups although the procedures we developed were somewhat different; the "concrete connection" was between items to be remembered and four chairs, with which the objects were said to "belong."

The first pilot experiment using chairs was run with two groups of ten- to fourteen-year-old schoolchildren in grades two to six. There were ten subjects in each group. The children all lived in Sergeant Kollie Town, a small roadside town located about two miles from Cuttington College. For both groups the experimenter stood behind four chairs, placed side by side, with a table on which the twenty objects were placed located behind him. The subject was seated facing the four chairs. The instructions used in earlier studies when objects were presented for recall were carried over unchanged to the present experiment, but the procedure was changed. On each trial when the experimenter held up an item to be remembered, he held it up over a particular chair for about two seconds, rather than simply holding it up where the subject could see it.

The two groups differed with respect to the assignment of items to chairs. In the rule condition, all items from a particular semantic category were held up over a particular chair. For instance, it might be that the items file, hoe, knife, hammer, and cutlass were held up over the chair on the far left of the subject; headtie, trousers, shirt, singlet, and cap over the adjacent chair, and so on. In the random condition, the semantic categories were broken up. A set of five items was always held up over a particular chair, but the selection of items did not constitute a naturally occurring linguistic group. It should be emphasized that the subject was not required to recall which chair an item had been held up over (as was the case in the paired-associate experiment discussed earlier), but only what items had been held up.

It is clear from Figure 4–9 that when items were assigned to chairs at random, the pattern of learning was very much like that we observed when objects were presented in the earlier experiments; recall averaged about twelve items, and improvement is restricted to the difference between the first two trials. For the rule condition, learning is more rapid and continuous, closely resembling the American pattern.

The same pattern emerges from an analysis of the clustering scores (Figure 4–10). There was significant *negative* clustering for the random condition (suggesting that subjects were clustering by chairs rather than by semantic categories), but highly significant clustering which increased over trials in the rule condition.

When these data were in hand, we felt that we had at last begun to approach an understanding of the dynamics of Kpelle memory organization. The replication sought to extend the basic findings of the pilot study by investigating such questions as the following: Would there be enhanced recall if there was only a single chair (perhaps using the chairs

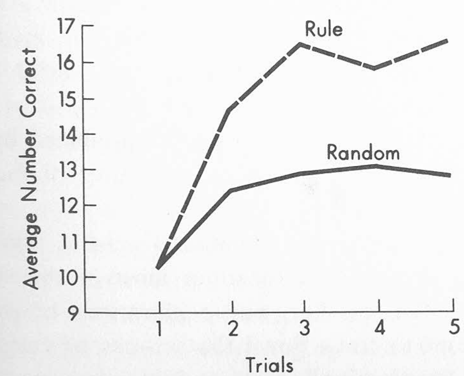

FIGURE 4–9 Average Number of Items Recalled per Trial for Rule and Random Groups in the Initial Study Using Chairs to Cue Recall

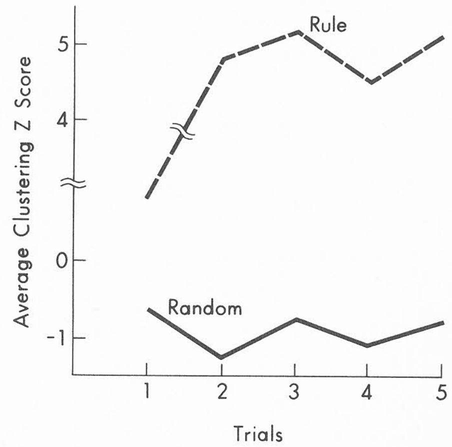

FIGURE 4–10 Average Clustering Scores as a Function of Trials in the Initial Study Using Chairs to Cue Recall

simply altered the subject's understanding of the task)? Would there be enhancement if items from a single category were always held up over the same chair, but the particular chair was changed from trial to trial?

To answer these and related questions, a large experiment was planned and executed. In view of the outcome, the strategy for conducting the replication could not have been conceived more poorly. A new experimenter was familiarized with the procedure. The experiment was conducted in a different area of Kpelleland than our previous studies, an area closer to the major population and rubber plantation center, and an area in which education was more widespread than it was around Cuttington College. The upshot of this experiment was that under *all* conditions, even one involving only a single chair, recall averaged seventeen to eighteen items per trial starting with Trial 1 and clustering was virtually perfect.

Further studies indicated that the effect produced by chairs as cues for recall depended heavily on specific features of the *experimenter's* behavior vis-à-vis the chairs.

However, this work was by no means a total loss; for one thing, *something* produced excellent recall in most of the studies using the chairs procedure. For another, our explorations of procedural variations in an attempt to track down the sources of variation go a fairly long way toward "explaining" (perhaps explaining away would be a better term) some of the gross differences in experimental outcomes.

We mention these experimenter differences in a series of experiments that we originally considered quite promising as an object lesson in the difficulties that can be encountered in research of any kind, but cross-cultural research in particular. The cases in which subjects' performance was all good or all bad speak only to the problem of experimenter differences. The cases where experimental treatments differ for the same experimenter indicate that genuine differential cuing effects can occur, but we do not know when they will affect both recall and organization and when they will affect only organization. (A general discussion of experimenter differences and other complications in conducting this kind of research are contained in Appendix F.)

Experiment 6: The Effect of Verbal Cuing on Recall

At the time of the first pilot study on the use of chairs as cues, when the subsequent experiments were not yet completed, we began to explore alternative ways of cuing recall. It seemed unlikely to us that cuing of the clumsy sort required by the use of chairs would be necessary to produce enhanced recall and organization. On both practical and theoretical grounds, we sought to determine if verbal cuing of any kind could influence recall in the way the chairs had in the initial pilot study.

Our first approach turned out to be much too subtle; we dropped one item from each category of the original clusterable list and inserted the category name instead. The results were identical to those obtained in the studies with the clusterable list and oral presentation.

Next we turned to a procedure based on the work of E. Tulving and his associates (Tulving and Pearlstone, 1966; Tulving and Osler, 1968). These experiments were concerned with the possibility that items might be remembered in the sense that they are stored in memory, but the subject cannot retrieve them at the time of recall. In Tulving's research, subjects were presented lists of words either with or without cue words present as a mnemonic aid. The cues were presented either at the time the to-be-recalled list was presented, or just prior to recall. It was concluded in the Tulving and Osler (1968) study that recall was enhanced only when cue words were present during learning, although it was considered possible that in the case of salient semantic groupings, the subject might provide the cue words (category names) for himself at the time of learning. The critical factor was thought to be the necessity for simultaneous storage of the item to be recalled and the retrieval cue.

Following the lead of Tulving and his associates, we sought to separate the use of cues during learning and during recall, and we varied the way in which these cues were presented: in some cases the subject was forced to recall by categories for part of the trials, in some cases this was optional.

The subjects were ten- to fourteen-year-old schoolchildren from the small road town of Wainsu, located approximately ten miles north of Cuttington College. The children were enrolled in grades two to six. Five groups of ten subjects each were run with the basic clusterable list for five trials. Presentation was oral.

The groups differed with respect to the point in each trial when category names were given as cues for remembering and the way in which recall of the items was elicited from the subjects. The basic four groups differed only with respect to when category cuing occurred; at the time the words were presented or at the time of recall. A two-by-two factorial design involving the four possible combinations of cuing at the time of presentation and recall represented the basic design. The instructions to each of these groups will make the group distinctions clearer:

1. Cued at both presentation and recall (cued-cued): "You and I are going to play. This play is about the things we work with. I will first call all of the names of these things. *The things will be clothing, tools, food, and utensils.* Listen to me carefully." (The items are then presented one at a time.) "Now I want you to name all of the things I told you; *they were clothes, tools, foods, and utensils.* When you are finished, tell me."
2. Cued at presentation, but not recall (cued-not cued): The instructions were the same except that the second italicized phrase was deleted.
3. Cued at the time of recall, but not during presentation (not cued-cued): The instructions were the same as the cued-cued condition, except that the first italicized phrase was deleted.
4. Never cued: Both italicized portions of the instructions to the cued-cued group were deleted.
5. Constrained recall: Subjects in this group were given the same instructions as the cued-cued group up to the point where recall began. Then for the first four trials, the subject was asked to recall the items by categories. For instance, he might be asked, "Tell me all the tools I named," then, "Now tell me all the foods I named," and so on. The order of recall categories was varied from trial to trial. When a subject had recalled all he could from a given category, the experimenter went on to the next until all four categories had been exhausted. This procedure was followed until Trial 5 at which time, without any warning, the procedure became that of the never cued group. That is, at the time of recall, the subject was simply told to name as many of the items told to him as possible.

RESULTS

In terms of the number of items recalled per trial, serial-position responding, and semantic clustering, the results for Groups 1 to 4 are entirely consistent with the results of previous experiments in which no cuing procedures were involved. There were no significant differences among the groups (which averaged about 11.5 items recalled per trial), very little improvement over trials, and a low level of clustering ($z = .49$). Although the absolute level of performance was slightly higher than that found in Experiment 1 where the same list and oral presentation were used, the level is typical of the experimenter for comparable experiments and of the same pattern and order magnitude as we found in Experiment 1. It would seem that giving subjects the category names as cues is not effective as a means for enhancing recall among our Kpelle subjects.

However, the results from the constrained-recall group indicate that verbal cuing *can* be effective under the proper conditions. Recall on Trial 1 was 16.6. On Trial 4 recall had improved slightly to an average of 17.6. Then, on Trial 5, when recall was completely unconstrained and uncued, recall for this group dropped only slightly to a level of 15.4, significantly better than the recall of the four other groups in the experiment (average = 12.9).

During the first four trials it was of course impossible to measure clustering for the constrained-recall group, which clustered perfectly by definition. However, on Trial 5 subjects in this group were free to recall items in any order they choose; clustering was the highest observed under any conditions in any of our previous experiments among the Kpelle ($z = 2.23$), except when chairs were used.

Thus, it would seem that a procedure that *forces* the use of cues has two effects on Kpelle free recall. First, it greatly increases the number of items recalled while the constraints are in effect. Second, it produces enhanced recall and semantic organization even after the constraints are lifted. How long such an effect would be preserved under unconstrained-recall conditions remains a topic for future research, but it seems that we have hit upon a way of "teaching" people to memorize more effectively in this short-term, free-recall task.

As interesting as the results from the constrained-recall group are, the question remains as to why the cued groups among the first four groups in this experiment failed to show any effects of cuing. The procedures we adopted were intended to be analogous to those used in similar cuing experiments in the Western psychological literature, but the outcomes were anything but similar.

A possible answer to this question may be contained in an examination of the adequacy of our "analogous" procedures. A re-examination of the work of Tulving and his colleagues indicated that their cuing was accomplished by providing the subject with a written list of the cue words during the appropriate cuing periods. Clearly, a procedure that relies on written words cannot be used with illiterates, but it is possible to use the procedures we developed in Liberia with literate American subjects. Although a comparison of Kpelle and American subjects using our procedure will not serve to specify the proper analogy between cuing procedures, it would serve the purpose of helping us to evaluate the effects of this cuing experiment. Consequently, a replication of this experiment was carried out with American subjects.

For our purposes here, the American replication of this experiment can be briefly presented. The subjects were third and sixth graders from the Laguna Beach, California, school system. The items to be recalled were the same as those used in the previously reported experiments where clusterable lists were employed with American subjects. Presentation and recall were both oral. The instructions were slightly elaborated versions of the instructions used for the comparable Kpelle groups. The experiment was conducted by Helen Wildwood.

In terms of the number of items recalled per trial and serial-position phenomena, the results obtained with our American schoolchildren were completely comparable to the results obtained with American schoolchildren reported on pp. 123–124. No significant effects were produced by the four cuing procedures used in Groups 1 to 4; a marked improvement took place over trials; a pronounced serial-position effect, dominated by recency was evident; and, a reliable difference between age groups appeared (the sixth graders recalled an average of 10.4 items per trial; the third graders, 8.2).

When we shift our attention to clustering effects, we find that overall there is no effect of the cuing conditions, but that late in training, the oldest subjects begin to show greater clustering in the cued-cued condition.

Most important, the American schoolchildren benefited from the highly constrained procedure of Group 5 in a manner quite similar to that observed with the Kpelle subjects. Recall was enhanced from the outset. However, in this case clustering, but not recall, was enhanced on

Trial 5, indicating that the conditions controlling transfer from the constrained- to unconstrained-recall conditions need further study.

Experiment 7: The Effect of Embedding To-Be-Recalled Material in Natural Verbal Contexts

We next wanted to determine if there was some experimentally accessible, but natural, situation in which rule-governed retrieval processes would routinely be used by Kpelle subjects.

In most of the previous research that we know of, the paradigm for the study of memory in naturalistic situations involved recall of stories. The classic research in this area is described in Bartlett's (1932) monograph to which we made reference earlier in this chapter. However, to embark upon a study of recall of stories would fail to tell us how the free-recall situations we have been studying make contact with normally occurring recall for connected material. Consequently, we have chosen a middle course, which, we think, permits us to link recall for connected and disconnected material. The basic strategy that we adopted was to provide a continuum of story contexts in which to present the twenty basic clusterable items used in most of the previous studies. These contexts varied from no context at all (our basic oral-presentation procedure) through a highly constrained story context in which each item was meaningfully linked to the neighboring item within the story. The subject was told a story and then asked to recall the items that figured in the story and in one case the story itself (see Appendix G for details). Responses were recorded and analyzed just as they had been in the previous experiments.

In one story a young man comes to the chief of a town and asks to marry the chief's daughter. He brings good bride wealth and the chief gives his daughter to the man. However, she soon discovers that he is a witch, and she wants to let her parents know where the man has taken her. So she leaves clues along the path as she travels to the man's farm. The clues (items) and their place on the path to the man's home make up the bulk of the story.

A second story involves four men who came to the town to ask for the chief's daughter. The first man brings five items of clothing; the second brings five items of food; the third brings five utensils; the fourth brings five tools. Once the story is told, the subject is asked what gifts were brought for the girl and then asked which man should get her.

The other stories contain the items in different arrangements designed to elicit differently structured recall.

In analyzing the results, it was found that the structure of the subject's recall mirrored the way in which the to-be-recalled items were structured within the story. If the items were structured in a linear manner (as in the first story described above), then a very high correlation between input and output orders was observed. However, if presentation structure was clustered, so was the structure of recall.

Discussion

Although by no means complete, the series of experiments presented in this chapter provides a wide range of situations in which to observe the influence of taxonomic categories on learning. And certain general patterns seem to emerge from the mosaic of results.

First, it is clear that under some circumstances all of the Kpelle groups studied were influenced in their learning by the presence of semantically definable categories. However, semantic control is neither uniform across groups nor across situations.

With respect to the various Kpelle groups studied, it appears from the evidence of our free-recall studies that semantic control (as manifested in clustering) became general in people with more than four to six years of schooling. How much more schooling we are not sure. Such control was not an inevitable consequence of maturation, because our adult populations differed little from younger groups.

The evidence from our American work shows really sizable amounts of clustering beginning to appear around the sixth grade, and it is possible that a similar finding could be obtained with the proper observations in Liberia. From Liberia we have the following pieces of evidence:

1. Comparisons of educated and nonliterate groups when the grades involved were second through fourth produced minimal clustering and slight differences in recall.
2. Comparisons of high-school students with nonliterate groups indicated rapid learning and significant clustering in the former.

Thus, it would appear that some time in the fifth- to eighth-grade range in Liberia, there is a change to a general use of semantic categories to control learning. We take it as no coincidence that these results parallel our findings in the similarity-mediation study in Chapter 3; the

conditions that produce taxonomic "sorters" are most probably related to the conditions that produce taxonomic "rememberers."

The lack of a general influence of semantic participation in the learning process among those who have not had extensive schooling raises the question of what general statement can be made regarding the conditions under which such participation is observed. We offer the following formulation: the conditions in which the influence of semantic categories and rapid learning are observed are those in which the situation is structured for the subject. The conditions of structuring need not make explicit the categories (as when a single chair is presented, or the subject sorts items into cups without categorizing them), but when the structure of the categories is made explicit (verbal discrimination, the chairs, constrained recall), strong semantic involvement ensues. When the structure is strong and anticategorical (as in the case of certain of our story-recall situations), that structure will dominate.

This pattern of results is strikingly consistent with the account of the development of recall performance offered by J. H. Flavell and his associates. For example, Moely, Olson, Halwes, and Flavell (1969) in discussing the results of their study remark:

> Both research findings and common sense lead one to suppose, for instance, that if a child of any age knows the name of an object he is instructed to remember, and if it also occurs to him to rehearse that name . . . then that rehearsal is very likely to help him remember the object. The problematical element in such a situation is precisely whether it will occur to him to rehearse, or for such situations generally perhaps, to engage in *planful* symbolic activity that is oriented towards and adapted to subsequent goal responses. [P. 32]

This description comes very close to characterizing the pattern of results from our Kpelle groups. The high-school student does not require specially structured situations in order to "have it occur to him" to use the semantic characteristics of the material to organize his recall—he produces that structure for himself. The nonliterate (and the same applies to those with little schooling) has not learned to spontaneously produce such structures under as wide a set of circumstances. He naturally uses them in some situations (when remembering stories) and can use them in a large variety of specially contrived situations (such as those provided by certain of our experiments).

In a number of instances we seemed to have tapped special organizing processes that permit the subject to retrieve the material he has been presented from memory. We have indicated certain naturally occurring situations in which such organization occurs spontaneously. We have laid to rest oversimplified ideas of rote memory and concrete mentality. But we have only nudged the iceberg that represents a full account of the processes underlying efficient and flexible learning (memorizing) and the cultural factors on which they depend.

We can look back on the verbal-discrimination studies as providing sufficient structure to induce the use of semantic information. Similarly, the chairs, constrained verbal cuing, and story experiments each, in its own way, provided such structure. But many of the simple free-recall studies did not, and hence learning failed to reflect the semantic structure of the to-be-learned material.

FIVE : Classification and Learning of Physical Attributes

"Why did you choose that one?" "Because it was beautiful."
ANONYMOUS KPELLE SUBJECT

Especially to those unfamiliar with the history of psychology, it may seem strange that the vast majority of American studies of classification learning (called, among other things, concept learning and discrimination learning) do not involve natural-language classifications of the things of experience. Instead, psychologists have relied heavily on the study of classification based on *physical attributes* (color, form, size, number, and so forth), which characterize some aspect of the things of experience.

For example, L. S. Vygotskii (1962) in his classic studies of the kinds of concepts formed by children of various ages, used blocks that differed from each other in height, width, shape, and color. The blocks belonged to different groups (determined by Vygotskii), which were signaled by arbitrary labels which served as a sign of category membership. By carefully noting the way in which his subjects tried to form groups during their search for the "correct" way, Vygotskii was able to establish developmental trends which he then related to his theory of cognitive development. The details of Vygotskii's experiment are not essential to this discussion, but the idea that "a concept" represents the combination of certain values (wide, blue, etc.) of the attributes (size, color, etc.) that are used to describe the set of blocks represents a very basic tool of Western developmental psychology.

Why this emphasis on artificial materials? Although it might be possible to make a case for the general importance of color, form, and similar physical attributes as the basis for classifications in Western societies (traffic signs, signals of various sorts), it could hardly be argued that such categories play as important a role in everyday behavior (Bruner, Goodnow and Austin [1956] give excellent intuitive accounts of such processes). It is much less reasonable to assume similarity of processes when our subjects are traditional Kpelle rice farmers. For the American adult, the use of conventionalized symbols to "stand for" other situations is commonplace; not only his written language, but his whole education, are based upon such symbolic activities. The Kpelle, too, makes use of symbols, but not symbols of this kind and not for this purpose. Consequently, we are going to have to be especially concerned with the relation between performance in our experiment and any underlying processes that we want to deal with in a speculative or theoretical manner.

A somewhat more subtle factor which enters into these experiments is that they almost all involve the use of materials that have no special meaning to the subject (blocks of wood or abstract designs on cards). This, of course, is a deliberate part of their design. But consider for a moment how rare a straight line, a perfect circle, or a pure, saturated red or green are in nature? In the Kpelle culture, which has no written language and only rudimentary pictorial art (such as patterns of cloth or painting on buildings), the standard classification of objects according to geometric form or pure, saturated color is not only a rare event, but probably contrary to experience.

A further problem concerns the relation between classification of the stimuli and the way in which the Kpelle language codes the particular dimensions involved. In the Kpelle language terms naming dimensions or attributes of experience are of several types, some of which seem to be different from English attribute names.

Number is one of the dimensions involved in our sorting problems, and our analysis of the language shows that numerical attribution is formed in a way very similar to that used in the English language (Gay and Cole, 1967; Gay and Welmers, 1970).

There are relatively fewer adjectives in Kpelle than in English. Of these, only the adjective translated "big" is a root word, while all the other adjectives are related to corresponding verbs. In particular, the adjectives naming colors are all related to verbs. There are three basic colors—white, red, and black. Each of these colors has a range of hue and saturation corresponding to it. Things are most frequently called white when they are of other colors, but are very low in saturation. The color red includes what in English would be named red, orange, dark

yellow, and even certain shades of purple. Objects are identified as black when they are of high saturation; the color black includes what would be called green, blue, purple, brown, or black in English.

Adjectives naming geometric shapes are named in a variety of ways, but rarely in the same ways as in English. Adjectives such as *round, square,* and *triangular* simply do not exist in Kpelle. There are free nouns which name certain objects, and which refer by extension to those shapes found in other objects. One term applies to pot, pan, frog, sledgehammer, and turtle and indicates circularity. Another term indicates triangularity and is used for a tortoise shell, arrowhead, bird's nest and bow. The term for a path refers equally to a straight and a curved line.

There is also a set of adverbs that suggest various complex textures and shapes. These adverbs are translated by such English adjectives as *smooth, crumbly,* and *jagged.*

This brief survey of terminology naming the attributes and dimensions present in our experiments can perhaps set a context for interpreting certain of the results. Number is named in much the same way as in English. Color is named by fewer words than is true in English, and they have definite, high-salience conventional meanings in terms of the quality of experience. (For example, white is said to indicate generosity and friendship, black shows evil and the intent to humiliate, and red suggests both ripeness and foreboding.) It may be that behavior toward colors in our studies is influenced by these attributes, rather than by hue and saturation, but we have no specific evidence on this point. Form also is described in different words, in different word-classes, and with different referents than in English. Here, too, some of the behavior of our Kpelle subjects with respect to form may ultimately be shown to depend on such differences in usage.

The general point to keep in mind is that the American emphasis on classification according to physical attributes, such as color and form, and the dependence of our research techniques on pictorial representation and nonmeaningful stimuli play directly to an area of experience almost wholly lacking among the Kpelle, and one in which particular language differences may play a role.

As a consequence, the results of any such studies will be especially difficult to interpret. These experiments violate the principles of research that we have used in the previous chapters of this book, and it might well be objected that we have effectively ruled out this kind of research as a vehicle for cross-cultural comparisons. We have resisted such a conclusion, although we realize that by adopting traditional psychological methods for the study of concept learning, we are exposing ourselves to a series of difficulties in the interpretation of data. It is our belief that if we are careful in the way that we evaluate exactly what it is that our subjects do when we present them with a classification task and if we restrict ourselves to inferences warranted from the data, artificially constructed experimental tasks can be useful in cross-cultural research.

The situations we have chosen for formal experimental study are all relatively simple. They were not chosen as representative of common Kpelle problem-solving situations, but rather as possibly useful special situations in which the general processes underlying problem solving and concept formation could be manifested clearly enough to permit detailed analysis.

We will begin by reraising the issue of how stimuli are classified. Various techniques for classifying artificial stimuli commonly used by psychologists will be discussed. Among these are *sorting, matching,* and *discrimination-learning* procedures. Both sorting and matching provide evidence about the physical attributes that are likely to be the basis of classification. The discrimination-learning studies address themselves to a variety of problems. Prominent in our work are the following questions: how does ease of learning depend on the particular stimulus attributes comprising the problem? Under what conditions will a learned classification transfer to new problems?

All of our studies yield data about the influence of schooling on basic classification-learning skills.

A major distinction to grow out of the research reported in this chapter is between two ways of learning to classify a set of stimuli that differ in principle. The first of these learning processes we characterize as *stimulus-specific* or *isolated*. By this we mean that the subject, when he chooses (say) a red triangle instead of a blue triangle, does so because he learned that the specific red triangle in question was correct on a previous trial. The second kind of learning we characterize as *general* or *concept-based*. By this we mean that in choosing the red triangle, the subject is basing his choice on knowledge that red pictures are correct, although he may or may not have learned that the particular red triangle he chose was correct previously.

This distinction, which we first encountered in the early sections of Chapter 4, turns out in the present context to have broad implications for the course of learning across a series of similar problems. Because

we think the issues raised by this series of studies are important for understanding cultural differences in cognition, we will present the course of our analysis in some detail.

Attribute Sorting: The Classification of Artificial Stimuli

Just as we began our study of cognitive processes operating on the domain of everyday objects with an inquiry into the way in which the objects were classified, so we will introduce our study of concept learning of artificially constructed stimuli with a study of their classification.

Dimensional Preferences and Free Classification

Our earliest foray into the study of how the Kpelle classify artificially constructed simuli was reported in our first monograph on the Kpelle (Gay and Cole, 1967). Groups of subjects (six- to eight-year-old nonliterates, ten- to fourteen-year-old literates, nonliterate adults) were presented eight cards and asked to sort them into two groups so that the members of each group seemed to go together. The cards differed in the color (red or green) of the stimuli pasted on them, the form of the stimuli (triangles or squares), and their number (two or five stimuli per card); color, form, or number could serve as the basis for forming the two groups.

We found that there were no striking differences among groups when *first* asked to sort these cards. Color and number were selected as the basis for sorting somewhat more often than form, but the difference was not statistically reliable. A second study, in which the forms were traditional, stylized human figures (male and female), confirmed the results of the initial study.

A significant difference among the groups appeared when subjects were asked to *re-sort* the cards and to find a second way to form groups. When we first tried this task, we found that few of our subjects would arrive at a new principle for grouping the cards once they had hit upon one of the three possibilities. After several abortive attempts at making our instructions more explicit, we finally constructed a completely different problem, and the subject was given an elaborate demonstration of the way in which the task was expected to proceed (at this point in our research we were more intent on obtaining an idea of the order in which Kpelle subjects would consider various attributes than in classification and reclassification skills per se).

Even with this elaborate demonstration procedure as a preliminary instruction, only the schoolchildren were generally able to come up with an alternative grouping of the cards, and it took them almost a minute and a half to do so. The two nonliterate groups experienced a great deal of difficulty and only about a half of each of these groups arrived at a second grouping. Most schoolchildren did a second sort, but were generally unable to find the third regular grouping (only 36 percent did so); among the nonliterate groups the third sort was very rare.

The same general findings were obtained by M. H. Irwin and D. H. McLaughlin (1970), who carried out a set of similar observations with members of the Mano tribe, neighbors of the Kpelle in Liberia. One of the useful additions provided by Irwin and McLaughlin was a sorting problem that was analogous to the Gay and Cole problem, but that used rice as the material to be sorted. Working with this traditional, extensively measured, and familiar material, Mano adults were better able to find new ways to form groupings; at least some of the difficulty in reclassifying the pictorial material was apparently caused by the material itself and was not a general inability to reclassify.

Dimensional Preferences Measured by Other Techniques

Before discussing the implications of our initial study of classification, the results of two additional studies intended to elicit information about how various stimulus dimensions are classified need to be considered. In the first of these studies (presented in more detail in Appendix H) a matching procedure was used. The subject was shown cards containing pictures of three figures. He was asked which two of the figures belonged together. Individual cards were arranged to permit matches based on color (red, white, and black), form (triangles and squares), and size (small or large). In this study form was used as the basis of matching far more often than color or size; color dominated only if the choice was between color and size alone.

In the second study subjects were presented with two cards, each with a stimulus set printed on it. For example, the two cards might depict a single red triangle and two blue squares. The subject was instructed to name one of the cards so that the experimenter would know which card the subject had in mind. A set of twelve such pairs was constructed and

presented to three groups of twelve subjects. The groups were six- to eight-year-old nonliterates, twelve- to fourteen-year-old schoolchildren (in grades four to six), and twelve- to fourteen-year-old nonliterates.

Among all groups the predominant response was for the subject to mention *only color* (for example, "It's the red one") in his response. Every subject mentioned color on every trial. Among the two nonliterate groups only two subjects mentioned number as well as color ("It's the two red ones") and only one mentioned form. Among the schoolchildren, five subjects mentioned number in addition to color, and only one mentioned form.

We present these data because they contrast so strongly with the results of our previous classification studies using stimuli of this type; color was not only dominant, it appeared to be the *only* dimension responded to by many of the subjects in the second study.

As a final complication these same stimuli were used in a sorting experiment analogous to that used by Gay and Cole (1967), and the results again indicated that color was far and away the dominant mode of classification.

Dimensional Preference?

If nothing else, this small set of studies of how the Kpelle will classify artificially constructed stimuli has amply demonstrated the validity of our warning that such material will present serious difficulties of interpretation. At the same time, the diversity of results with procedures that seem to be similar (and that generally produce similar results when conducted on subjects from Western countries) raises a number of issues that are central to our concern and about which we can offer a little evidence from the experiments presented in this chapter.

One conclusion seems fairly safe at this point: we are not going to be able to draw general inferences about the developmental significance of some particular order of classification (such as color over form) until we can specify the rules which lead one order to dominate under some conditions and another order to dominate under other conditions. In this respect we simply have to advise caution, although inferences based on dimensional preferences are widely used in cross-cultural research (Bruner, Olver, and Greenfield, 1966; Serpell, 1969). Second, any study of learning based on stimuli of this type must take account of the particular way in which the specific stimuli used are likely to be classified. It is not possible to make safe a priori assumption about color or form classification in general.

Problems in Re-Sorting and the Transfer of a Discrimination

One of the more striking results from our earlier sorting experiments, as well as from those of Irwin and McLaughlin (1970), was the difficulty that our subjects, particularly our nonliterate subjects, experienced when asked to find a basis for sorting other than their original basis. This failure to reclassify is especially unusual when it is remembered that they had been given a demonstration problem in which the possibility of alternate modes of classification had been pointed out. Although striking, the difficulty in reclassification is very difficult to interpret. Even though we provided an example problem to our subjects, there was some question about how they interpreted the instruction to regroup a set of pictures which they had already been told was "fine." Even if there was no ambiguity about the instructions, and the difficulties with the second sort were the result of some deeper kind of misunderstanding, we cannot be certain what the extent of the problem was. For example, could subjects who experienced difficulty in spontaneously re-sorting these pictures learn to do so if a trial-by-trial *learning* procedure was used?

In order to answer this question we conducted a rather extensive experiment with five- to seven-year-old children who had not, for the most part, attended school, although some had entered the first grade to learn English. The stimuli were the same cards used in our original sorting experiment; they contained red or green figures that were triangles or squares, with two or five figures on each card. The procedure was quite different from that employed in the sorting studies. A subject was not presented the set of cards all at once, but rather was asked to discriminate between cards in terms of specific dimensions when the experimenter held up a pair of cards that differed along all three dimensions simultaneously (for example, two red triangles versus five green squares). The subject was instructed as follows:

> I will show you two papers. Each time I show you these papers I want you to tell me which one I am thinking of. You must give me the one I am thinking of. If you are correct, I will say yes. If you are wrong, I will say no. You must try to be correct every time.

Each subject was presented a total of three problems. Every possible combination, including repetitions, of dimensions was sampled. When a dimension served as the basis for solution more than once, correct value was reversed between problems. For example, some subjects received a problem set where color was relevant on all three problems. In this case a particular subject might have the sequence: red correct, green correct, red correct. Other examples of problem sequences are color-color-form and color-form-number. This procedure yielded twenty-seven possible groups. Eight subjects were run in each subgroup for a total of 216 subjects. This extraordinarily large number of possibilities and subjects was included because we wanted to sample a large number of conditions under which subjects were required to switch the basis of their classification. Subjects continued responding until they were correct nine trials in a row or until forty trials had been presented.

On the first problem slightly more than half of the children were able to solve the form problem; in contrast, the color and number problems were solved by almost everyone. Ignoring the particular sequence of problem types, there was virtually no improvement for later problems in the three problem series. Only if the subject was presented a homogenous series of problems (for example color-color-color) was there any indication that learning on later problems was faster than learning on the initial problem. These results are graphically presented in Figure 5–1. On the left-hand side of the figure, we see the average number of trials to criterion (a score of forty was applied if criterion was not met) for homogenous problems. On the right-hand side of the figure, we see the results for completely heterogeneous problems (for example, color on problem two had been preceded by a form or number problem, and so forth). Even this rather crude representation of the data clearly indicates that there was a general absence of improvement unless the same dimensions served as the basis of solution throughout the experiment. Even under optimal conditions, there was very little improvement for the form problems.

Verbal justifications offered by the subjects, as we had come to expect from our earlier verbal-discrimination experiments, were very rare. However, they were qualitatively in line with the learning results. Only two subjects could verbalize the basis for a form solution—both said that they chose the "houses" (triangles). Twenty-two subjects identified

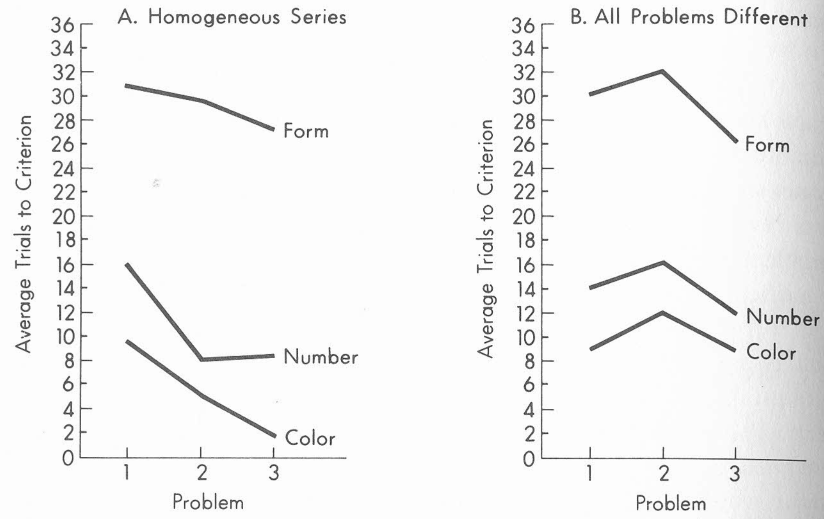

FIGURE 5–1 The Number of Errors Committed When Subjects Are Presented Problems Based on Color, Form, or Number. (In A, subjects have a single dimension relevant throughout training; in B, the dimension is changed from problem to problem.)

number solutions; in half the cases the number two was named and in most of the remaining cases the subject used the term *many*. Thirty-six subjects identified the color solution, using the term *black* for green and *red* for red.

These results suggest, first of all, that the failure to spontaneously sort along the form dimension persists even when an explicit training procedure is used. Before jumping to the conclusion that there is some general disability in understanding of form classes among this subject population, we need to consider additional facts. In the first place, there is no question that these children can make form discriminations—if they could not do so, they would be unable to function as human beings. The question is whether or not they can learn form classes with *stimuli such as those we have used here*. We know that under some conditions, at least, they can learn form discriminations of the kind presented here. In an investigation of discrimination among elementary geometrical-form classes (Gay and Cole, 1967, pp. 54ff.), children in the same general age range learned to discriminate the *class* of triangles from the *class* of squares in about thirteen trials.

But in the earlier work there were *no alternative dimensions on which the subject could base his response;* only the forms varied among

the stimuli. This suggests that it may be competition among dimensions that is causing trouble for these subjects. A second suggestion from this study is that interproblem improvement in learning rate ("learning to learn") will occur only if the particular dimension serving as a basis for solution is maintained from one problem to the next. In order to clarify these issues, we need different, and, in general, simpler experimental situations, which will permit us to disentangle the possible processes involved.

Basic Procedure: Discrimination Learning and Transfer

As an example of one specialized approach, consider the following situation. The subject is shown two stimulus blocks, one a large black square, the other a small white square. These stimuli differ from one another on two dimensions; black versus white and large versus small. In a simple discrimination training procedure, this pair of stimuli is presented and the choice of one of the blocks is rewarded, that is, is said to be correct (for instance, the large black square). In this example the basis for the discrimination used by the subject is uncertain. Assuming that the blocks are presented on the left- and the right-hand of the subject in a random order, there still remains an important source of ambiguity. When the large black square is identified as correct, is it correct because of its largeness or its blackness? With the use of only one pair of stimuli, it is impossible to clarify this ambiguity. Therefore, in a typical experiment of this type a second pair of stimuli is presented on half of the trials, interspersed with trials on the first pair. This second pair performs the function of breaking up the ambiguity. In the particular case in point, one would introduce a large white square paired with a small black square. By evaluating the pattern of correct and incorrect choices on the two pairs, the subject can infer the dimensional basis for discrimination. If the experimenter designates (reinforces) both the large black square and the small black square on successive presentations, the subject could infer that "black" is correct. If, however, the experimenter reinforces the large black square and the large white square, the subject could infer that "large" is correct.

After the subject has reached some reasonable criterion of successive correct responses, there is still some ambiguity, but in this case the ambiguity rests in our inference about the basis for his selections. We might want to conclude that he has correctly analyzed the pattern of reinforcements and has acquired the concept in dimensional terms. For example, he may say to himself "black is correct" or "large is correct," or more complexly, "brightness is the relevant dimension, and black is the correct value of that dimension." However, there is another possibility: the subject may have acquired two separate choice responses, "large black square" and "small black square." These responses would imply that the two pairs of stimuli in the acquisition phase were learned as independent discriminations, and did not have a dimensional basis that bound them together.

One way to decide among these two possibilities (dimensional learning that includes both pairs of squares or learning of two independent discriminations) is to design transfer studies that indicate how the original discrimination was learned. The transfer conditions most widely applied for this purpose have been termed a "reversal shift" and "nonreversal shift" (Kendler and Kendler, 1967). A reversal shift is defined by a within-dimension change in reinforcement patterns (see Figure 5–2). For example, if a subject has been trained on "black" (either large or small square) we may shift our reinforcement to "white" (either large or small square). This change remains within the color dimension. What this means for the subject is that for *both* pairs in the original training, he must now shift his choice from the block that was previously correct to the previously incorrect block. A nonreversal shift, by contrast, involves a shift of reinforcement to a dimension that had not been previously employed, for example, a shift from "black" to "large." Considering block pairs instead of dimensions, this means that the subject must relearn only *one* of the previously learned discriminations. For example, if he had previously learned that the large black and a small black squares were correct, he can maintain his response to the small black square when shown small-black–large-white and must only change his response to choose a small white square when shown small-white–large-black. The subject must maintain his response on one discrimination pair but shift the response for the other. These relationships are shown schematically in Figure 5–2.

This analysis of reversal and nonreversal shifts would seem to indicate a clear advantage for transfer to a nonreversal condition if the subject is treating pairs separately, since he must only relearn one of the discrimination pairs. In the reversal-shift condition, he must relearn both. This pair-by-pair analysis, however, assumes that the two discriminations can be treated independently (by the experimenter in analyzing shifts in reinforcement and by the subject in relearning the discrimination). There is considerable evidence to suggest that rats (Kelleher,

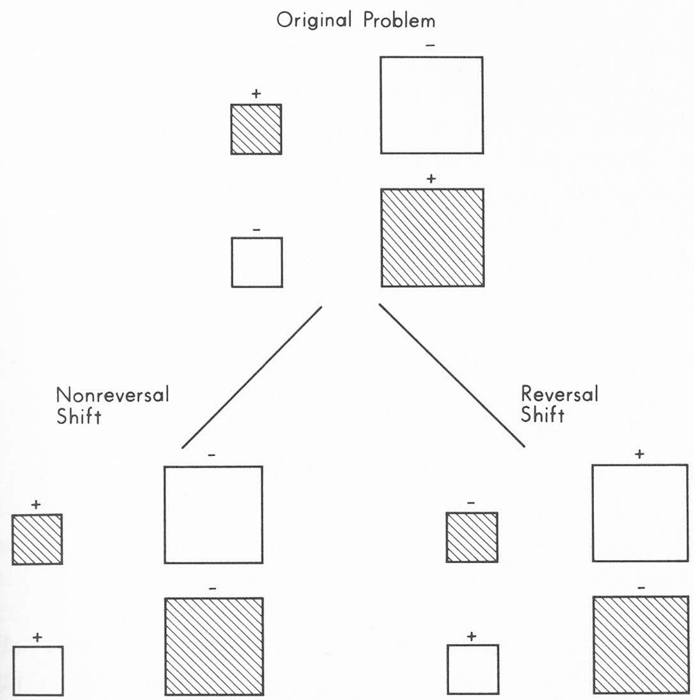

FIGURE 5–2 Schematic Representation of Discrimination-Learning Experiment in Which the Basis of Solution Is Shifted Following Initial Learning

1955) and young human children (Kendler, Kendler, and Wells, 1960) perform in exactly this manner.

However, our intuitive description of this problem is in terms of stimulus dimensions and attributes ("the black ones are correct"), and the pairs are treated as instances. What if subjects are, indeed, responding in terms of our casual description and using the stimulus dimensions to bind together their responses to the two subproblems? In this case making a double change might be easier than changing only one discrimination. We may, therefore, argue that reversal shift should be easier than nonreversal shift for people who approach the problem in this way, since a reversal allows the subject to make a *patterned* change to "the problem" as a whole.

Pursuing this line of argument, we may say that, if it is shown that reversal shifts are learned more rapidly than nonreversal shifts, the reasonable conclusion is that the subject is responding to the two pairs of stimuli in a common fashion; in contemporary terminology, he is using the stimulus dimensions to mediate his performance. The bulk of the evidence indicates that older subjects (beyond kindergarten) make reversal shifts faster than they make nonreversal shifts, lending support to the analysis presented here. (In somewhat different terms, this analysis and the data upon which it is based are presented in Kendler and Kendler, 1968.)

Recent analyses of the discrimination-transfer process, while supporting the general developmental trend that we have just described, suggest that the learning processes involved may be analyzable into a combination of elementary processes, which include, but are not exhausted by, our description of "typical" adult performance given above. As we shall see, we believe this analysis to be directly relevant to an understanding of cross-cultural differences in elementary concept learning.

The Pseudoreversal Paradigm: Stimulus Dimensions versus Response Patterning

Consider again the reversal-nonreversal paradigm diagramed in Figure 5–2. In light of our previous discussion, we can see that two things are involved. On the one hand, there are two pairs of stimuli to be learned before we introduce the reversal of the subject's response to either one pair (nonreversal shift) or two pairs (reversal shift) of stimuli. The presence of multiple stimulus dimensions presumably influences the transfer (shift) learning. However, it is possible to learn this problem without using dimensional information by learning that two independent discriminations are presented, and continuing to respond in the successful manner until a clue is given that something about the problem has changed (for example, the previously correct member of a pair is now incorrect). A subject who learned in this manner might reason, "if this first pair has been changed, perhaps the other changed too." Further, he might do all of this without ever saying to himself that the two discriminations involve the same stimulus dimension—he may only be picking up response patterns.

The possibility of such pattern learning in young American subjects was investigated by B. Sanders (1971). Her experiment was exactly analogous to the reversal-nonreversal paradigm illustrated earlier, with the important exception that the two discriminations to be learned bore no obvious dimensional relationship to one another. One pair of stimuli consisted of two crosses—one red and the other black. The other pair

consisted of a square and a triangle, both of which were gray. This arrangement makes a dimensional response to the first pair ("the red one is correct") of no help in learning the second discrimination ("the square one is correct").

With a two-component problem of this type, it is possible to investigate the effects of response patterning in shift behavior independent of the stimulus dimensions. If the subproblems are learned independently, we ought to be able to see this independence operate during the shift phase of the experiment. Continuing the analogy with the typical discrimination-shift experiment, Sanders changed the reinforcement contingency on one or both of the discriminations. Among younger children she found that changing only one pair (pseudononreversal) was an easier transfer problem than changing both pairs (pseudoreversal). These results parallel results obtained when dimensions are common to both stimulus pairs. Sanders found that the superiority of nonreversal among younger children was the result of their tendency to treat the discriminations independently. However, dependence between the two subproblems characterized the learning of the older subjects. That is, once the shift phase began, a nonreinforcement on one subproblem immediately led to a spontaneous change of choice on the second subproblem. This spontaneous shifting resulted in errors if the second pair was not changed (nonreversal) but immediate solution if the second pair *was* changed (reversal).

This evidence strongly suggests that in addition to (and even possibly instead of) any dimension-related mediation of responses involved in discrimination transfer, there are strategies of response-patterning which operate in this situation to unite subproblems. As the most elementary form of discrimination-reversal problem, we will begin our discussion of discrimination transfer among the Kpelle here.

Methods and Procedures

Two pairs of stimuli were used in the acquisition and shift phases of the study. Pair 1 consisted of two T-shaped cutouts of three-eighth-inch plywood of equal size; one was red and the other black. Pair 2 was a triangle and square of roughly equal size cut out of three-eighth-inch plywood. The triangle and square were left unpainted (and, hence, highly similar in color). The pairs were presented repeatedly throughout the acquisition series in a randomized order, which varied both the suc-