JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 6, 212–230 (1968)

Concept Formation by Kindergarten Children in a Card-Sorting Task[1]

PATRICK SUPPES AND IRENE ROSENTHAL-HILL[2]
Stanford University

Fifty children in kindergarten learned to sort cards according to one of four attributes on each of three different dimensions. Geometrical cards had the dimensions shape, border and color, and "people" cards depicted men, women, and children, wearing clothes of various colors and engaging in different activities. For the first two sets of problems, information was presented to the subject by two positive examples. Half of the sample had geometric problems first while the other half began with "people" problems. For both groups positive transfer occurred from the first set of problems to the second set. Problems with "people" cards were more difficult than problems with geometrical cards. After reaching criterion on the first two sets of problems, subjects quickly mastered problems with a positive example and a negative one. An all-or-none model was found to be inadequate as a description of the learning process. Subjects achieved few correct solutions, however, before reaching criterion and correct responses tended to occur in the last phase of the experiment.

Experiments on concept formation by children have been derived from a variety of theoretical orientations and have used different types of experimental conditions. Since many of the experiments have not used careful controls and have not included a detailed analysis of the learning processes, only a somewhat vague picture of the process of concept formation has emerged from the research.

Experiments have shown that concepts become more difficult to learn as they increasingly depend on abstract processes rather than processes directly related to perception (cf. Inhelder and Piaget, 1964; Osler and Fivel, 1961; Osler and Weiss, 1962). An increase in the tendency to use mediating concepts rather than to acquire direct stimulus–response connections has been found with increasing age in experiments by Kendler (1963, 1964) and by Youniss and Furth (1964).

---

[1] This research has been supported by the United States Office of Education under contract 3-10-009 with Stanford University.

[2] We are grateful to Mr. Victor Norton, the principal of the elementary school from which subjects were drawn, and his teachers for their cooperation.

When a conflict was introduced between perception and logical processes that were to be performed, the concepts became especially difficult for young children (cf. Smedslund, 1961a, 1961b, 1962, 1963a, 1963b, 1963c; Beilin, 1964; Halpern, 1965). Attempts to teach children under age 10 concepts for which abstract thinking must be dominant over perception have been generally unsuccessful in the relatively short periods of time used in the experiments.

The construction of detailed mathematical models to describe the process of learning concepts has been the aim in research with children conducted by Suppes and his collaborators (Suppes and Ginsberg, 1963; Suppes, 1965b, 1966). The simplest model, the all-or-none model, assumes that a child moves in one trial from an unconditioned state in which he makes correct responses at random to a conditioned state in which he always makes a correct response. Most of the data obtained by Suppes and his collaborators could be fitted by an all-or-none model of learning, but in the last quartile of responses prior to attaining criterion, there was usually some improvement in performance. Conceptually, however, the all-or-none model has more severe limitations. It does not adequately describe the process by which the child acquires a new concept, or transfers concept learning to a new stimulus situation. The present paper is part of a series of investigations aimed at providing a deeper analysis of the process of concept acquisition.

The experiments to be described in this paper studied the learning of concepts by children in kindergarten. Concepts which were to be identified involved geometrical dimensions (e.g., shape) and dimensions related to people (e.g., activity). Responses were recorded throughout the experiment to permit a detailed analysis of the process. Transfer effects were studied by having the children perform the concept-identification tasks in different orders.

The experiments, which are described in completely self-contained fashion here, follow the general sort of procedure used by Suppes and Schlag-Rey (1965). They are part of an extensive series of experiments that attempt to get a better grip on the processes of concept formation by requiring more complicated classifying responses. The underlying assumption is that by asking the *S* to classify a number of stimulus items at the same time, possibly the entire population of stimuli, a much clearer and explicit indication of the underlying hypotheses he is using in making the classification is given. When the *S* is asked to classify the entire population of stimuli, the most natural theoretical assumption is that he exhibits the first-order hypotheses he is using in making the responses themselves. It is a natural additional theoretical assumption to suppose that in back of the first-order hypotheses more general second-order hypotheses are rules (e.g., in the present experiments the first-order hypothesis of a simple sort, in the case of the geometrical cards would be that the concept is that of square, and the appropriate second-order hypothesis that would be used to reach this first-order hypothesis would be the relevant dimension of shape).

Within this general theoretical framework, a primary objective of the present experiment is to explore the range of validity and limitations of an all-or-none learning model for responses of the degree of complexity used here. None of the experiments reported in Suppes (1965b) or Suppes and Ginsberg (1963) involve anything but classifying responses to single stimulus items.

METHOD

Stimulus Materials

Two sets of 3 × 5 cards were constructed. Each set consisted of cards varying in three four-valued dimensions. One set of cards, the geometrical cards, had the dimensions of shape, color, and border. The shapes were squares, triangles, circles, and diamonds. The colors which filled the area on the cards between the shapes and the borders were red, blue, green, and yellow. The borders surrounding the edge of the cards were striped, dotted, plain, or filled with black ink. The other set of cards, the "people" cards, had the dimensions of types of people, activities, and articles of clothing which were also distinguished by color. The types of people were a boy, a girl, a man, and a woman. The people were engaged in activities of sitting, running, walking, or standing. The types of colored clothing were yellow hats, blue coats, red scarves, and green jackets. Each of the two sets consisted of 64 cards formed from all possible combinations of the four attributes of the three dimensions.

Experimental Procedure

The problems in the experiment were to be solved by grouping cards according to one attribute of one dimension. The information necessary to solve the problems in the main part of the experiment was presented by two positive examples for one set of problems, and a positive example and a negative one for the other set. The two examples were chosen so that only one classification using one attribute of one dimension was possible (e.g., the positive cards and the negative cards for a problem differed in the attribute of the relevant dimension, but had the same attributes on the other two dimensions). The two methods of giving information were used with problems from both geometrical and "people" cards to give a total of four different sets of problems. In each set there were 12 different problems with each having one of the four attributes of one of the three dimensions as an answer. Six problems used a set of ten cards with four cards (including the displayed cards) falling into one class, and the remaining six problems used 11 cards with five cards comprising a class.

The sortings were performed on a table which was divided into halves by a strip of tape. At the top of each half of the table, a small area was marked off with tape where the cards serving as examples of the concept were to be placed. All the cards for each problem were given to the child at once. The child was always told to place all cards that were similar to the positive cards with these cards, and the remainder on the other side of the table. For the problems for which two positive cards were shown, *E* stated that the two cards which he was putting on the same side of the table went together. When *E* gave a negative stimulus card, he said that this card was different from the card that was already on the table and would, therefore, be placed on the other side of the table. Then, *E* instructed the child to put the cards that were the same as the positive stimulus card with this card, and the cards that were different on the other side of the table with the card that was displayed there. During the first three experimental sessions, the preliminary and control sessions, the child was not told whether he was sorting the cards correctly. Subsequently in the learning sessions, the child was shown the correct solution for every problem. Testing was done on consecutive days unless illness or vacations interfered.

Preliminary session. In the first session of the experiment, subjects were given three problems to familiarize them with the experimental procedure. For each problem, one card was placed on one side of the table, and the subjects were told to place with this card all other similar cards. A total of five cards with pictures of animals was used for the first problem, and for the other two problems, 12 geometrical or 12 "people" cards were used.

Control sessions. In the second and third sessions, the subjects were given the first two or three problems from each of the four sets that were to be given during the main part of the experiment. In addition, two geometrical problems were given which had two positive examples, and which could be solved by using the two dimensions conjunctively. The control sessions provided a baseline against which to measure learning that occurred in the experiment.

Learning sessions. In the fourth session of the experiment, subjects began learning a set of problems with two positive examples shown. Half of the subjects (Group 1) had geometrical problems first while the other subjects (Group 2) began with "people" problems. After reaching criterion on their first set of problems, the children were given the other set of problems with two positive examples (e.g., children in Group 1 transferred to "people" problems with two positive examples after they had reached criterion on geometric problems). Children who had reached criterion on these first two sets of problems were given geometric problems with a positive example and a negative one, and after they had mastered these they proceeded with a set of "people" problems which had a positive example and a negative example.

The following procedure was followed for each of the four sets of problems:

(a) In the initial sessions, six problems were given each day with two problems drawn from each dimension. On the first day for a problem set, two attributes of each dimension provided solutions for the problems and the next day the remaining attributes of each dimension were used. The two sets of attributes for each dimension continued to be covered on alternate days for each problem set until four out of five consecutive problems had been solved correctly for one dimension. (Problems given on two or three consecutive days had to be considered in determining whether the criterion had been attained.)

(b) After criterion had been reached on a dimension, problems from this dimension were discontinued. Thus, only four problems (covering the other two dimensions) were given each day after criterion had been reached on one dimension. After criterion had been reached on two dimensions, only two problems were given each day.

(c) After criterion had been reached on each dimension separately (as described above), six problems drawn from the three dimensions were again given as had been done in the beginning of the series of experimental sessions. In this phase of the experiment six problems were given each day until all problems in one session had been solved correctly.

Modifications in the procedure were introduced if the subject had not reached criterion by the eighth session. At the beginning of the ninth and tenth sessions of problems with two positive examples, subjects were asked to describe one card. If they did not spontaneously mention the attributes of the three dimensions, they were told them. In the eleventh session and all subsequent sessions of all sets of problems, the subjects were asked to give the basis for the correct sorting, and if they did not respond correctly, they were told the answers. (In the first ten sessions, the correct sorting had been displayed while in the latter sessions verbalizations were also supplied, e.g., subjects were told, "These are all squares"). The maximum number of sessions for any subject was 37.

Subjects

Twenty-one boys and 29 girls in kindergarten were tested. The average age of the subjects during the experiment was 5½ years (*SD* = 3 months). Twenty-five children completed all four sets of experimental problems. The remaining 25 children had reached different stages of the experiment when the school year ended in June. Most of the subjects who did not complete all parts of the experiment were those who were tested in the latter part of the school year when there was not enough time left to complete the experiment. Forty-three subjects reached criterion on at least their first set of problems, and three additional subjects reached criterion on the three dimensions of the set separately. All 50 subjects contributed data for the analysis of responses in the preliminary and in the control sessions. For the learning part of the experiment the maximum number of subjects contributing data was 25 for Group 1, and 21 for Group 2.

RESULTS

Method of Analysis of Results

Classification of responses to problems. Responses to problems were classified into the categories *one-dimensional responses, disjunctive responses,* and *conjunctive responses.* For one-dimensional responses, all cards in one class had the same attribute on one dimension. Disjunctive responses consisted of a set of cards which had either of the two attributes of one dimension which occurred on the two positive stimulus cards. For conjunctive responses, all cards had the same attribute on each of two dimensions. All responses that did not fall into any of these classifications were designated unclassified.

Analysis of the learning process. Indices of learning were computed for the different sets of problems and were compared to determine whether some types of problems were easier than others, and whether there were transfer effects from one set of problems to another. Measures of performance included the number of sessions required to reach criterion, and the average number of problems solved before criterion was reached. A more detailed analysis of the learning process was made by studying sequences of responses. Tests of the all-or-none model of learning were made by using chi-square tests for order in pairs of adjacent problems, and for number of correct responses in blocks of three problems. Stationarity was tested in backward- and forward-learning curves and in Vincent curves. The statistical tests have been described in Suppes and Ginsberg (1963). For all statistical tests in this paper, the .05 significance level was used.

CLASSIFICATION OF RESPONSES

Sortings in the initial sessions. Responses to all the problems in the first three sessions, when no answers were given, were analyzed together. Classifications of the responses were based on dimension and on type of response. The latter classification consisted of the three types of responses described previously, and the fourth additional category, *incorrect one-dimensional response.* The latter group included responses where all the cards placed with the positive stimulus card had the same attribute on one dimension, but the negative stimulus card which the subject had ignored showed that the response was incorrect.

The only dimension that was used frequently in the initial sessions was that of shape. In the nine geometrical problems, shape was used in an average of 5.4 problems ($SD = 2.4$). The other two dimensions of the geometrical cards, color and border, were used much less frequently as a basis for sorting, and there was no tendency to use these dimensions throughout the nine problems. In six "people" problems, the most popular dimension, activity, was used more than once by only 18% of the *S*s and was used once by 58% of the *S*s. In the first of the "people" problems, where one card was displayed, sex was used the most rather than activity. For geometrical problems, shape was the most common dimension at the beginning of the experiment as well as in the second and third sessions.

The *S*s used all the types of responses which enabled them to sort by shape throughout the problems. Correct one-dimensional shape responses could occur on the preliminary problem where only one card was displayed, on one of the problems where a positive example and a negative one were given, and on three problems which had two positive examples presented. For two of the latter problems a conjunctive classification was also possible, but was used by only six *S*s for the first problem, and two *S*s for the second problem. Disjunctive shape responses and incorrect one-dimensional shape responses could each occur on two problems. Nineteen *S*s used disjunctive shape responses for the two problems, and 16 *S*s used this sorting for one problem. The incorrect one-dimensional shape response was used by 19 *S*s for the two problems and by 13 *S*s for one problem.

Few *S*s were able to use the negative information which was given in three geometrical problems and two "people" problems in the control sessions. The *S*s could achieve one correct solution to these problems if they had a tendency to use the dimension which happened to be correct for the problem, but to have more than one correct answer they had to use the information from the negative example. For the "people" problems, only three *S*s obtained two correct answers, and for the geometrical problems nine *S*s had two correct sortings. (Only one *S* had the three geometrical problems correct.)

CLASSIFICATION OF RESPONSES TO PROBLEMS IN THE MAIN PART OF THE EXPERIMENT

Responses to problems in the main part of the experiment in which two positive instances of the concept were presented to the *S* were classified according to the previously described categories of one-dimensional, disjunctive or conjunctive response. Responses to all problems for each dimension prior to the criterion success run of problems were included in the tabulations. The proportion of classified responses ranged from approximately one-fifth to one-third for the different sections of the experiment. Most of the classified responses were correct one-dimensional responses, and the other classifications were rarely used. Thus, *S*s learned quickly that disjunctive shape responses, which they had used frequently in the initial experimental sessions, were incorrect.

To determine whether the low proportion of classified responses was due to *S*s not examining all the cards carefully to check whether a card had been overlooked, responses with one card missing from the group of cards placed with the displayed cards were included with the classified responses instead of with the unclassified responses. With this modification in the method of classification, the proportion of responses falling into some classification was raised somewhat and fell in the range of approximately one-third to one-half in different portions of the experiment.

LEARNING CORRECT SOLUTIONS TO PROBLEMS

Number of Sessions to Reach Criterion

Problems with two positive instances. Averages for the number of sessions to reach the various criteria are shown in Table 1. With both geometrical and "people" cards, the group that was given the problems as their second set of problems was superior in performance. The differences between the two groups in the number of sessions to reach the various criteria were significant when a t-test was applied with the degrees of freedom calculated by Welch's method to take account of the unequal variances. (The values of t ranged from 2.85 to 5.90 and df had a minimum value of 13.) After completing "people" problems, *S*s quickly reached criterion on geometrical problems. The *S*s who began the experiment with geometrical problems still required a number of sessions to master the "people" problems, but their performance was superior to the group that began the experiment with these problems. This difference in performance on geometrical and "people" problems was also found for the sets of problems given first to *S*s; fewer sessions were required to reach the criteria for geometrical problems than for "people" problems.

As shown under the heading "all" in Table 1, *S*s quickly learned to solve problems of all types in a single session after they had reached criterion on the individual dimensions (The heading "all" here refers to the criterion of five out of six problems' being correct). Thus, although many *S*s had repeated exposure to one or two of the dimensions with the remaining one or two dimensions omitted, they quickly regained their proficiency in solving problems from the dimensions that had been learned the fastest. Also, *S*s had little difficulty in shifting between dimensions on successive problems in the last portion of the experiment when they were again given all types of problems in one session. Attainment of the more stringent criterion of all problems' being correct in one

TABLE 1
NUMBER OF SESSIONS TO REACH CRITERION[a] ON THE PROBLEMS WITH TWO POSITIVE INSTANCES PRESENTED

	Group 1 (geometrical problems first) (23 *Ss*)				Group 2 ("people" problems first) (18 *Ss*)			
	Geometrical problems							
No. of sessions	Shape	Border	Color	All	Shape	Border	Color	All
Mean	6.3	10.5	10.5	13.1	2.4	3.3	2.9	4.5
SD Mean	4.6	5.6	6.5	7.0	0.7	1.5	1.3	1.9
Median	5	12	10	15	2	3	2–3	4
Range	2–16	2–23	2–31	2–35	2–4	2–7	2–7	2–9
	"People" problems							
No. of sessions	Activity	Color	People	All	Activity	Color	People	All
Mean	6.5	4.7	9.7	11.9	12.7	10.6	14.2	16.7
SD Mean	4.2	3.8	5.3	6.1	6.3	4.9	4.1	4.9
Median	5	3	9	14	13–14	11	14–15	17
Range	2–17	2–14	2–18	2–25	2–24	3–20	4–22	5–25

Note.—Two problems from each dimension given in each session.

[a] To reach criterion on each separate dimension *S*s had to solve four out of five consecutive problems. To reach criterion on all dimensions *S*s had to solve five out of six problems in a single session after (or simultaneously with) reaching criterion on all separate dimensions.

session occurred very quickly after the criterion of five out of six problems' being correct had been reached. More than half the *S*s reached the two criteria simultaneously on both sets of problems, and more than two sessions were required by only three *S*s for the "people" cards, and by no *S*s for the geometrical cards. (The maximum number of sessions was four for "people" cards).

Problems with a positive and a negative instance presented. The *S*s reached criterion very quickly on the problems for which a positive example and negative example of the concept were presented after criterion had been reached on problems with two positive examples. Results for the two groups of *S*s were very similar and the statistics in Table 2 are for the combined groups. On each dimension half of the *S*s reached criterion without an error. After reaching criterion on the separate dimensions, *S*s quickly demonstrated their ability to solve problems from all dimensions in one session. All *S*s, except one, reached the criterion of all six problems' being correct in one session at the same time as they reached the less stringent criterion of five out of six problems' being correct. The results which have been discussed and which are shown in Table 2 include only those *S*s who had begun the experiment early in the school year and had reached this part of the experiment. (Thirteen *S*s from each group participated in this experiment).

TABLE 2
NUMBER OF SESSIONS TO REACH CRITERION ON PROBLEMS WITH A POSITIVE AND A NEGATIVE INSTANCE PRESENTED

No. of sessions	Geometrical problems (26 *S*s)			
	Shape	Border	Color	All
Mean	3.6	3.2	2.8	4.2
SD Mean	2.6	1.9	1.5	2.9
Median	2	2	2	3
Range	2–11	2–9	2–6	2–10

No. of sessions	"People" problems (25 *S*s)			
	Activity	Color	People	All
Mean	3.2	3.5	2.8	4.6
SD Mean	2.8	3.2	1.3	3.9
Median	2	2	2	4
Range	2–16	2–18	2–7	2–20

Note.—Two problems from each dimension were given in each session.

EFFECT OF MODIFICATIONS IN PROCEDURE IN LATTER PART OF THE EXPERIMENT

At the beginning of the ninth and tenth sessions, *S*s identified the attributes of the three dimensions on a given card. In the eleventh session and in all subsequent sessions, *S*s were told the attribute that was relevant for the classification for each problem whenever they did not verbalize correctly (e.g., when the correct answer was displayed in the eleventh session, the child was asked, "Why are these cards the same?" and if the answer was incorrect, he would be told, for example, "These are all circles"). In order to find out whether these modifications in procedure were affecting performance, two measures of performance, proportion of correct solutions and number of *S*s reaching criterion, were compared before and after the new procedure had been introduced. Only a few sets of data were available for these analyses since the *S*s had reached criterion before the modifications were introduced in most parts of the experiment.

The proportion of problems solved was unaffected by the procedural modifications. The number of *S*s reaching criterion also remained unchanged in sessions 10 and 11, the sessions which would show the effect of identifying the three attributes, but there was an increase in sessions 13–16, the sessions which would show the effect of the verbalization procedure. (For the comparisons the same number of sessions was used preceding the sessions mentioned as was used to measure the effects of the modifications).

Analysis of Learning Curves

A number of statistical tests were applied to the data to determine whether an all-or-none model of learning was applicable. Since the model postulates that the probability of a correct response (guessing probability) remains constant for each trial prior to the last error, tests were applied to those responses which occurred before the last error had been made. Data from individual *S*s were averaged in several different ways to obtain learning curves. Forward-learning curves that were constructed by averaging responses for *S*s beginning with the first trial permitted study of the initial phase of learning. To study the portion of learning near the point where criterion was reached (this point occurred at different trials for different *S*s) backward curves were used starting with the trial before the last error and going backwards. Also, Vincent curves which combined responses occurring in each quartile of trials for each subject were constructed to show whether probability of success remained constant in comparable phases of the experiment for each *S*.

Geometrical problems. Only the data from *S*s of Group 1 were analyzed since the *S*s of Group 2 reached criterion too quickly on the geometrical problems. Table 3 shows the proportion of correct responses and the average number of correct responses for each dimension prior to the criterion success run. The *S*s had a low proportion of correct responses to problems on the border and color dimensions before the criterion success run, but they had a somewhat higher proportion of success on the shape dimension. Fewer *S*s contributed data for the shape dimension than for the other two dimensions. An especially high rate of success was obtained for the first problem of the experiment which was solved by 13 of the 15 *S*s. (The two displayed cards showed circles).

TABLE 3

GROUP-1 PERFORMANCE ON PROBLEMS PRIOR TO REACHING CRITERION RUN FOR GEOMETRICAL PROBLEMS WITH TWO POSITIVE INSTANCES PRESENTED

Measures of performance	Shape[a] (15 *S*s)	Border (23 *S*s)	Color (24 *S*s)
Number of problems	178	401	403
Problem of last error			
Mean	11.9	17.4	16.8
SD	9.1	10.4	12.2
Number of successes			
Mean	2.5	1.1	1.0
SD	1.3	1.9	1.4
Average proportion success	.23	.07	.07

[a] Geometrical problems were presented first.

With respect to the Vincent curves, which are shown in Fig. 1, chi-

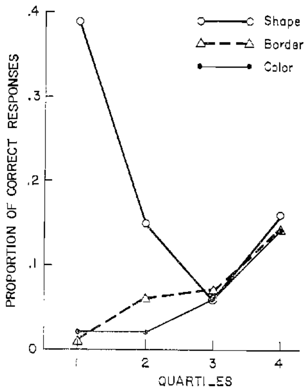

FIG. 1. Vincent learning curves in quartiles on geometrical problems of Group 1 for proportion of correct responses prior to the last error.

square tests rejected the hypothesis of constant proportion of success in the quartiles for each dimension. (The chi-squares were 12.4 for shape, 11.1 for border, and 13.5 for color, with $df = 3$). For the border and color dimensions, the highest proportion of success was obtained in the last quartile, but for the shape dimension the largest proportion occurred in the first quartile. The number of *S*s contributing data for the Vincent curves was ten for shape, 19 for border, and 21 for color, and the average number of problems per quartile was 2.2, 3.8, and 3.6 for the three dimensions, respectively.

The hypothesis of stationarity was also rejected for the backward-learning curves of the border and color dimensions, which are shown in Fig. 2. (The chi-squares were 13.2 for shape, 22.5 for border, and 37.0 for color, with $df = 10$). Forward curves were stationary for the border and color dimensions, but not for the shape dimension. At the cut-off point for these learning curves, four *S*s remained for the shape dimension, and seven for the other two dimensions.

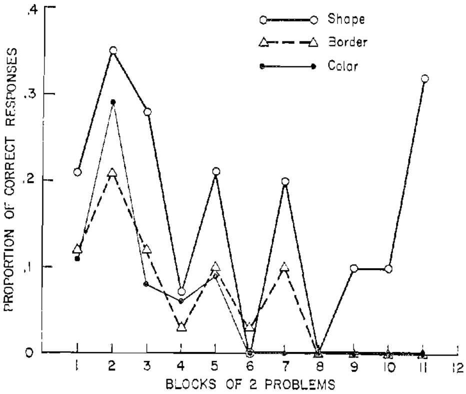

FIG. 2. Backward-learning curves on geometrical problems of Group 1 for proportion of correct responses prior to the last error.

"People" problems. As shown in Table 4, *S*s who had been given geometrical problems before the "people" problems achieved more correct solutions to problems from the activity and color dimensions than the *S*s who began the experiment with "people" problems. Only some of the *S*s of Group 1, however, showed superior performance on these two dimensions. Thus, when the proportions were recalculated by first computing the proportions separately for each *S* and then averaging, a range of .1–.2 was obtained for the proportions.[3] Two of the problems from the

---

[3] Proportions computed with subjects weighted equally had approximately the same values as those computed with all responses combined for all sets of data except those mentioned above.

TABLE 4
PERFORMANCE ON PROBLEMS PRIOR TO REACHING CRITERION RUN FOR "PEOPLE" PROBLEMS WITH TWO POSITIVE INSTANCES PRESENTED

	Group 1 (Geometrical problems first)			Group 2 ("People" problems first)		
Measure of Performance	Activity (17 *S*s)	Color (15 *S*s)	People (21 *S*s)	Activity (19 *S*s)	Color (21 *S*s)	People (21 *S*s)
Number of problems	174	106	329	174	106	329
Trial last error						
Mean	10.2	7.1	15.7	23.2	17.0	25.8
SD	7.9	7.9	9.7	9.7	10.9	10.5
Number of successes						
Mean	3.0	2.2	2.0	4.1	2.5	1.5
SD	3.0	3.0	2.5	3.8	2.5	1.9
Average proportion success	.32	.36	.14	.18	.16	.06

activity dimension had a higher proportion of success than the other two problems. (The problem for which the answer was *sitting* had a higher proportion of success in both groups, and the problem for which the answer was *walking* had an elevated rate of success for *S*s of Group 1.)

Chi-square tests applied to the Vincent curves, shown in Fig. 3, rejected the hypothesis of constant proportion of success throughout the

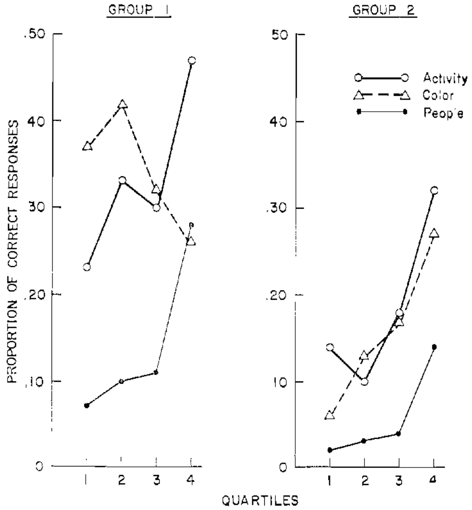

FIG. 3. Vincent learning curves in quartiles on "people" problems for proportion of correct responses prior to the last error.

quartiles for all dimensions for subjects of Group 2 as had been the case with the geometrical problems of Group 1 ($\chi^2 = 17.1$ for the activity dimension, $\chi^2 = 13.0$ for the color dimension, and $\chi^2 = 21.1$ for the people-type dimension, with $df = 3$). The average number of problems per quartile was 5.1 for the activity dimension, 3.7 for the color dimension, 5.9 for the people-type dimension with 17 *S*s contributing data for the first two dimensions, and 21 *S*s for the third. The hypothesis of stationarity was also rejected for the people-type dimension for Group 1 ($\chi^2 = 16.0$, $df = 13$), the only dimension which had sufficient data for Group 1 to give a meaningful statistical test. Eighteen *S*s from Group 1, who had an average of 3.4 problems per quartile, contributed data for the people-type dimension. For the activity and color dimensions, the number of *S*s for whom data were available was 11 and 7, and the average number of problems per quartile was 1.8 and 1.3, respectively. In all the nonstationary curves, the rise in proportion of success occurred in the last quartile.

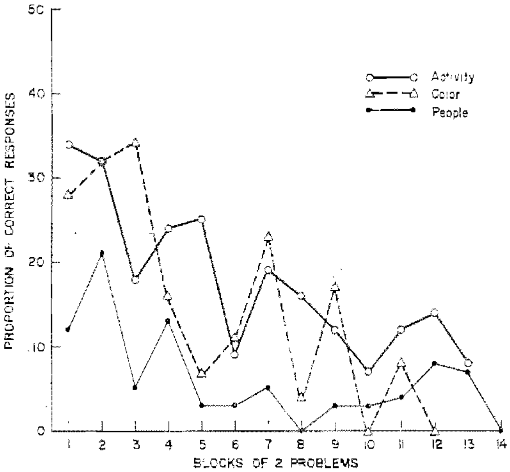

FIG. 4. Backward-learning curves on "people" problems of Group 2 for proportion of correct responses prior to the last error.

In the backward-learning curves for Group 2, shown in Fig. 4, the hypothesis of stationarity was rejected for the color dimension ($\chi^2 = 26.9$, $df = 11$) and the people-type dimension ($\chi^2 = 27.5$, $df = 13$). There was no consistent decline in performance on successive problems in these curves, but there was considerable fluctuation in proportion of success. An attempt was made to find an explanation for the fluctuations by investigating the effect of verbalization, and the effect of different rates of success for different problems of a single dimension, but neither of these explanations was compatible with the data. For the activity-dimension curve of Group 2, for which the chi-square value ($\chi^2 = 19.2$, $df = 12$) just missed reaching significance, there was a higher proportion of success in the two blocks of problems before the last error. For the backward-learning curves for Group 1, shown in Fig. 5, the hypothesis of stationarity was not rejected, although definite trends away from stationarity seem evident in the figure. All the forward-learning curves were stationary except the curve of Group 2 for the people-type dimension. At the cut-off point for the learning curves, the number of *S*s remaining from Group 1 was five for the activity dimension, six for the border dimension, and four for the color dimension, and the number of *S*s left from Group 2 was six for each dimension.

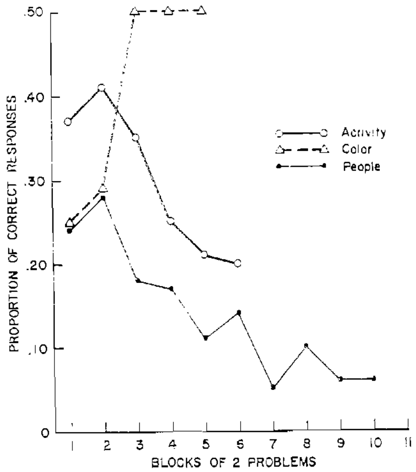

FIG. 5. Backward-learning curves on "people" problems of Group 1 for proportion of correct responses prior to the last error.

Two other tests of the all-or-none model of learning led to rejection of the model.[4] A test (the order test) was made of the hypothesis that the response to a problem was independent of the response made to the preceding problem. The second test compared the number of successes in blocks of three consecutive problems with the frequencies predicted from the binomial distribution. Only data from the activity and color dimensions were used for these tests since expected values for the third dimension fell below seven. Significant chi-square values were obtained for the order tests in all sets of data except the data of subjects of Group 2 from the activity dimension. The fit of the binomial distribution to the

---

[4] For the geometrical problems, these two statistical tests had to be discarded because expected values were too low.

number of successes in blocks of three problems was adequate for the data of *Ss* of Group 1, but not for the data from Group 2. Since the expected frequency for zero successes was very low, the category of zero successes was combined with the category of one success for all sets of data except the data of the activity dimension of Group 1 [e.g., the number of successes in blocks of three problems was (a) zero or one, (b) two, or (c) three].

DISCUSSION

In the present experiment, *Ss* learned the mediating concepts of dimension rather than specific solutions for all problems except for two problems from the activity dimension. Capability of children in kindergarten in using mediational processes has been demonstrated in Kendler's experiments (1963, 1964). In an experiment of Youniss and Furth (1964) children of age 10 learned specific responses rather than the general concepts of conjunction and disjunction, but children of age 13 learned the general concepts. The concepts which were studied by Youniss and Furth were not, however, as familiar to children as those in the present experiment.

In a recent article, Haygood and Bourne (1965) discussed the importance of considering both the recognition of relevant attributes and the rules utilized to form concepts, and they pointed out that most research has been concerned with the first of these factors. In the present experiment, acquisition of a method of utilizing the information provided by the two displayed cards was an important part of the learning process. Thus, after the *Ss* were told what the relevant attribute was for each problem, they still required additional sessions to learn to solve the problems. Also the fact that positive transfer was obtained between geometric problems with two positive examples and the corresponding "people" problems showed that the *Ss* had learned a general method of solving the problems since color was the only common dimension for the two types of cards. The group of *Ss* who began the experiment with a set of problems (either geometric problems or "people" problems) required more sessions to reach criterion than the group that received the problems as their second set.

Bower and Trabasso (1964) have developed a theory of concept identification which assumes that the learning process is all-or-none and relates the rate of learning to the probability that the *S* samples the relevant attribute. Some of the experiments to which Bower and Trabasso applied their theory used the same relevant dimensions of border and color that were used in the present experiments. However, their experiments were designed so that only the factor of perceptual identification of the relevant attributes was important in the learning process, and the responses required only classification of single stimulus items.

Application of the all-or-none model of learning to the data of the card-sorting experiments reported here yielded results similar to those obtained by Suppes (1965b) and Suppes and Ginsberg (1963), even though the responses required of *S*s were considerably more complex than what was required in the earlier experiments. In particular, most of the statistical tests applied to the data of the present experiment rejected the hypothesis that no learning had occurred before the trial of the last error, but the data rejecting the all-or-none model occurred almost entirely in a few trials before criterion. From a theoretical standpoint, these results add weight to the general conclusion that an all-or-none model provides a first approximation to response data from a wide variety of concept formation experiments with children, but a more complex model is needed to go beyond the first approximation.

A second more profound limitation of the all-or-none model is discussed in Suppes (1965a) and has been mentioned earlier in this paper. Roughly speaking, the idea is that the all-or-none model does not specify the mechanism which an *S* is using in forming a new concept. In the present experiments, for example, the ultimate selection and use of the relevant dimension by *S*s meeting criterion cannot be understood in any deep sense simply by considering the all-or-none model. Additional assumptions about hypothesis sampling or some similarly structured process are clearly needed. Steps in this direction for some closely related experiments are taken in some unpublished works of Suppes and Schlag-Rey. Their basic idea is to postulate first- and second-order hypothesis sampling with the appropriate a posteriori Bayesian distribution dominating the sampling distribution. As far as we could determine, the qualitative aspects of the present experiment are consistent with their theoretical ideas, but detailed application to the data reported here was not feasible because the expected values of many "cells" were too low.

REFERENCES

BEILIN, H. Perceptual-cognitive conflict in the development of an invariant area concept. *Journal of Experimental Child Psychology,* 1964, **1**, 208–266.

BOWER, G. H., AND TRABASSO, T. R. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford University Press, 1964, 32–94.

HALPERN, E. The effects of incompatibility between perception and logic in Piaget's stage of concrete operations. *Child Development,* 1965, **36**, 491–497.

HAYGOOD, R. C., AND BOURNE, L. E. Attribute- and role-learning aspects of conceptual behavior. *Psychological Review,* 1965, **72**, 175–195.

INHELDER, B., AND PIAGET, J. *The early growth of logic in the child.* New York: Harper, 1964.

KENDLER, T. S. Development of mediating responses in children. *Monographs of the Society of Research in Child Development*, 1963, **28**, 33–52.

KENDLER, T. S. Verbalization and optional reversal shifts among kindergarten children. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 428–436.

OSLER, S. F., AND FIVEL, M. W. Concept attainment. I. The role of age and IQ in concept attainment by induction. *Journal of Experimental Psychology*, 1961, **62**, 1–8.

OSLER, S. F., AND WEISS, S. R. Studies in concept attainment. III. Effect of instructions at two levels of intelligence. *Journal of Experimental Psychology*, 1962, **63**, 528–533.

SMEDSLUND, J. The acquisition of conservation of substance and weight in children. III. Extinction of conservation of weight acquired normally and by means of empirical controls on a balance. *Scandinavian Journal of Psychology*, 1961a, **2**, 85–87.

SMEDSLUND, J. The acquisition of conservation of substance and weight in children. V. Practice in conflict situations without external reinforcement. *Scandinavian Joural of Psychology*, 1961b, **2**, 156–160.

SMEDSLUND, J. The acquisition of conservation of substance and weight in children. VII. Conservation of discontinuous quantity and the operations of adding and taking away. *Scandinavian Journal of Psychology*, 1962, **3**, 67–77.

SMEDSLUND, J. Development of concrete transitivity of length in children. *Child Development*, 1963a **34**, 389–405.

SMEDSLUND, J. Patterns of experience and the acquisition of conservation of concrete transitivity of weight in 8-year old children. *Scandinavian Journal of Psychology*, 1963b, **4**, 251–256.

SMEDSLUND, J. Patterns of experience and the acquisition of conservation of length. *Scandinavian Journal of Psychology*, 1963c **4**, 257–264.

SUPPES, P. The kinematics and dynamics of concept formation. In Y. Bar-Hillel (Ed.), *Proceedings of the 1964 international congress for logic, methodology and the philosophy of science.* Amsterdam: North-Holland Publishing Co., 1965a, 405–414.

SUPPES, P. On the behavioral foundation of mathematical concepts. *Monographs of the Society of Research in Child Development*, 1965b, **30**, 60–96.

SUPPES, P. Mathematical concept formation in children. *American Psychologist*, 1966, **21**, 139–150.

SUPPES, P., AND GINSBERG, R. A fundamental property of all-or-none models, binomial distribution of responses prior to conditioning with application to concept formation in children. *Psychological Review*, 1963, **70**, 139–161.

SUPPES, P., AND SCHLAG-REY, M. Observable changes of hypotheses under positive reinforcement. *Science*, 1965, **148**, 661–662.

YOUNISS, J., AND FURTH, H. G. Attainment and transfer of logical connectives in children. *Journal of Educational Psychology*, 1964, **55**, 357-361.