MATHEMATICAL CONCEPT FORMATION IN CHILDREN

PATRICK SUPPES
Institute for Mathematical Studies in the Social Sciences, Stanford University

T
HE research reported here has been conducted over the past 4 years with a large number of younger collaborators. As is characteristic of much scientific work, the efforts we have made have not been concentrated on any single point of theory. We have attempted to penetrate the complex problems of mathematical concept formation at several points, but I do not think we have yet been very successful in formulating fundamental theory. For the present report I have selected certain experiments that bear upon fundamental issues that must be faced by any theory adequate to the facts of mathematics learning. Our reasons for running these experiments and designing them as we have are, of course, not restricted to the particular issues I shall discuss. Some of the experiments were in fact conceived in terms of probing the limits of mathematics learning in the classroom for elementary school children of various ages.

The issues that I want to discuss may be rather arbitrarily placed under four headings. I begin with a discussion of the evidence for all-or-none conditioning processes in simple concept learning. I next turn to questions of transfer. I then consider an experiment concerned with the geometric invariants of the perceptual space of children. Finally, I consider an investigation just recently completed on the mechanisms of concept formation. In this last experiment I hope we are beginning to dig deeper into the fundamental processes of concept acquisition.

ALL-OR-NONE CONDITIONING PROCESSES IN SIMPLE CONCEPT FORMATION

One of the most discussed issues in current psychological theory is the extent to which learning and retention occur on an all-or-none basis. An appraisal in terms of several types of experimental situations is to be found in Estes (1964) and Suppes and Ginsberg (1963). What I want to present here is evidence that the learning of simple mathematical concepts is more closely approximated by an all-or-none assumption than by any simple incremental assumption.

The all-or-none postulate directly leads to the following model, formulated in stimulus-response terms. The child begins the experiment by not knowing the association established by the experimenter between individual stimuli and responses. He is thus in the unconditioned state. On each trial there is a constant probability that he will pass from the unconditioned state to the conditioned state. While he is in the unconditioned state it is postulated that he will simply guess the correct response with probability \( p \), and this guessing probability is independent of the trial number and the preceding pattern of responses. When the child reaches the conditioned state the probability of making a correct response is postulated to be 1 for the remainder of the experiment. It should be apparent that this model postulates an all-or-none character for the learning process. This means that for an individual subject the theoretical individual learning curve has the simple appearance of Figure 1.

The important thing to note about Figure 1 is that the learning curve is perfectly flat until the last error occurs before the point of conditioning, and at that point there is a strong discontinuity. It is to be emphasized that, in contrast to the learning curve shown in Figure 1, when data are pooled over a group of subjects the usual in-
incremental learning curve is obtained. This is because individual subjects are becoming conditioned on different trials, and averaging across subjects produces the usual incremental curve.

There are, however, several simple ways of analyzing group data in order to see if the flat, discontinuous sort of curve shown in Figure 1 is obtained. (Ideally, of course, we would like to look at the learning curve for individual subjects, but in the learning of many concepts there are not sufficient trials to provide adequate data on individual subjects.) The essence of any one of several approaches is to look at response data only prior to the last error made by each subject. We may plot a backward learning curve from this point or, what is essentially an equivalent procedure, we may plot Vincent learning curves in percentiles of trials rather than simply trials prior to last error. If the all-or-none model is supported by the data these Vincent learning curves should look approximately like the curve shown in Figure 1.

In contrast, the simple incremental assumption is incorporated in a linear incremental model. This model postulates that the probability of making a correct response increases linearly each time the subject is exposed to the stimulus and is shown the correct response. The learning curve postulated for an individual subject by the linear model looks something like Figure 2. The mean learning curve for the linear model has the same shape as the individual learning curve, and is also not affected by restriction to data prior to the last error.

Keeping the two kinds of theoretical learning curves shown in Figures 1 and 2 in mind, let us now look at the Vincent learning curves prior to last error from three mathematical concept-formation experiments with children. Before turning to the first experiment, there are two remarks I want to make about the experiments reported in this paper. With one exception, the experimental procedure called for an immediate correction of a wrong response, that is, the child was shown immediately after making his own response the correct response, and in some cases was also required to make an overt correction response. Also characteristic of all the experiments is the fact that the stimulus display on a given trial was never repeated. Many of the classical concept-formation experiments in psychology actually use experimental procedures that repeat a stimulus display on several trials. When such a procedure is used, it is not a simple matter to decide whether the learning of the correct response is to be attributed to concept formation or to a simple stimulus-response association. By changing the stimulus on each trial this confounding is avoided, and no simple principles of stimulus-response association are adequate to explain the learning of the concept. It is necessary to postulate that the association is between concepts and responses, and not simply between stimuli and responses. Another way of putting the matter is that in order to explain learning in the kind of concept experiments we have performed, it is necessary to impose a well-defined structure on the set of stimuli, and not simply to take the stimuli as arbitrary members of an unstructured set.

The first experiment is concerned with the learning of a simple pattern. The problem was very close to that of the familiar oddity problem much studied with monkeys, but in this case the subjects were 7- to 8-year-old children in Brussels, Belgium, and the experiment was performed in collaboration with Rachel Joffre of the Laboratory of Psychology, University of Brussels. On each trial the child was shown three standard dot-dash patterns immediately above the response keys, as

![Stimulus](image)

**Response 1** Response 2 Response 3

*Fig. 3. Typical stimulus display in pattern experiment.*
indicated in Figure 3. At the top of the screen, well above the dot-dash patterns, the subject was shown three stimuli in a horizontal row, with two of the stimuli the same and one different. The task for the child was to match the position of the dot with the position of the odd stimulus and then to press the response key below the matching dot-dash pattern. Thus for the display of Figure 3, Response 3 is correct. The stimulus at the top of the screen changed on every trial, but the dot-dash patterns above response keys were the same on all trials. However, the response position of a given dot-dash pattern was randomized from trial to trial. This proved to be a difficult task for children of this age. A Vincent learning curve on data prior to last error in percent correct responses and mean latencies in seconds for pattern experiment \((N = 25)\) is shown in the lower half of Figure 4 for the 25 subjects who satisfied the strong criterion of 18 consecutively correct responses. There is slight evidence of an increase in the proportion of correct responses in the last two or three sextiles before criterion, but the relative flatness of the curve is much more favorable to the all-or-none than to the simple incremental model. Latencies for the six sextiles as well as the three corresponding sextiles extending beyond criterion are shown in the top half of Figure 4. The point to remark on here is the slight increase in latencies in the fifth and sixth sextiles just before criterion is reached. I shall briefly return to this point later.

The second experiment is concerned with the simplest sort of mathematical proofs. It was performed in collaboration with John M. Vickers. The mathematical systems used for the experiment are concerned with the production of finite strings of 1's and 0's. Any finite string of 1's and 0's is a formula of the systems. The single axiom is the single symbol 1. The System U4 has four rules of inference as shown in Figure 5. The letter \(S\) stands for any nonempty string. Thus the first rule \((R1)\) says that if we have a string we may add two 0's to the right of it. A typical theorem and its proof are shown at the right of Figure 5. I have chosen this particular theorem because it is the shortest one that uses all four rules of inference. Line 1 of the proof begins with the axiom. We apply Rule 1 to the axiom to obtain 100 and then eliminate the extra 0 by using Rule 3. Next we apply Rule 2 to obtain Line 4. We now have the theorem except for an excess 1. We finally apply Rule 4 to eliminate the extra 1. The System U5 is like System U4 except that U5 has an additional rule permitting a 1 to be removed from the left. This talk of strings, axioms, and theorems may seem rather fancy when I mention that the experiment was run with first graders. As you would expect, the language I have been using is not the language we used with them. Moreover, quite apart from questions of verbal description, we did not ask them to write down proofs in the fashion shown in the figure, but rather proceeded in the following fashion. A child was shown a horizontal panel of illuminated red and green squares. Below this panel was a second panel with matching squares. The first square in the lower panel was always illuminated red, corresponding to the single axiom, 1. The child was shown four buttons that he could use to light up additional squares or remove lighted squares from the lower panel, working from left to right. The changes in the lower panel made by pushing the four buttons corresponded precisely to the four rules of U4 (in the case of System U5 there were of course five buttons). A pictorial description of the result of punching each of the buttons was shown directly above the buttons. The child was told that his objective was to punch

**Mathematical Concept Formation**

<table>
<thead>
<tr>
<th>Rules of Inference</th>
<th>Theorem III</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R1). (S \rightarrow 500) ((1))</td>
<td>1</td>
</tr>
<tr>
<td>(R2). (S \rightarrow S1) ((2))</td>
<td>100</td>
</tr>
<tr>
<td>(R3). (50 \rightarrow S) ((3))</td>
<td>10</td>
</tr>
<tr>
<td>(R4). (S1 \rightarrow S) ((4))</td>
<td>101</td>
</tr>
<tr>
<td>((5))</td>
<td>10</td>
</tr>
</tbody>
</table>

**Fig. 5.** Rules of inference and typical theorem in proof experiment.
the buttons so as to make the row of colored lights in the lower panel match the row of colored lights in the upper panel. The green lights corresponded to 0's and the red lights to 1's. It should be apparent that the sequence of responses emitted by the child in attempting to match the upper panel is formally isomorphic to writing a proof in the style shown in Figure 5.

Ignoring certain minor complications, the stimulus discrimination that the child faced on each trial is simply described (for a theoretical discussion, see Suppes, 1965). He must compare his panel, the lower one, to the top one. This comparison will classify the stimulus discrimination into one of four categories: (a) additional green squares need to be added to match the theorem (R1); (b) additional red squares need to be added to match the theorem (R2); (c) a green square must be deleted to continue to match the theorem (R3); or finally (d) a red square must be deleted in order to match the theorem (R4).

Each subject was presented with 17 theorems per session for a total of 72 trials. The subjects consisted of 50 first graders; 25 were first run on System U4 and then on System US. This order was reversed for the other 25. In Figure 6 Vincent learning curves are shown in quartiles for trials prior to the first run of five successively correct responses. (A correct response is defined in the obvious way in terms of the four possible stimulus discriminations defined above.) The Vincent curve for System U4 is based on 572 observations and the Vincent curve for the more difficult System US on 1,360 observations. An incremental effect, particularly in the fourth quartile, is apparent in the data for US. In the case of U4 no incremental effects are noticeable. Other aspects of the data of U4 also fit very well the all-or-none model. For example, the binominal distribution of responses in blocks of four for the same portion of the data is shown in Figure 7. The number of possible errors in a block of four trials is on the abscissa and the frequency of these errors on the ordinate. As would be expected from the apparent fit, a chi-square test shows that the two curves are not significantly different.

Turning now to the third experiment, I would like to begin with a remark about its pedagogical origin. For the past several years we have been running a pedagogical experiment on the teaching of mathematical logic to able fifth and sixth graders. As is always the case with such pedagogical programs, we have from time to time been faced with the problem of what to do with new members of classes in which the whole class is part of the program. Some of the teachers have claimed to observe that the fastest way to catch up these late arrivals has been to teach them first the formal machinery of logic and afterward to give them the intended interpretation of the symbols, for example, the standard interpretation of the ampersand as "and" and the wedge as "or." The experiment I want to describe was designed to throw some light on the validity of this observation. The complete design of the experiment...
is rather complicated in terms of the choice and order of sentences and the measures used for judging their logical complexity. For purposes of the present discussion, however, the essential facts may be presented. The subjects were 48 fourth graders with an IQ range from 110 to 131. Twenty-four of the subjects began with the formal system and 24 began with the interpreted system. In the interpreted system only English words and sentences were used and the subject was asked to make logical inferences by selecting among four alternative answers the single one that was a logically valid conclusion from the given premises. The stimulus material given the formal group was logically isomorphic in every respect, but only uninterpreted letters and symbols, as shown in Figure 8, were used. The figure shows the three rules of inference presented to the formal group. The first rule on the left is the rule of detachment: from P and if P then Q, infer Q. The second rule, shown in the middle of the figure, is the rule of simplification: from P and Q we may infer P. The third rule is the commutative rule for "and" : from P and Q we may infer Q and P. I emphasize again that the formal group was not told the meaning of the arrow (implication) or of the upside-down wedge (conjunction).

On the first day of the experiment for both groups four trials were devoted to each of the three rules and then trials were given in blocks of six. Each subject was required to run at least three blocks, and the criterion of success was one block of six trials of entirely correct responses. The experiment was continued for a second day for those subjects who did not meet criterion on the first day, and similarly for a third day. If a subject did not meet criterion by the end of the third day he was switched in any case to the other system on day 4. The Vincent learning curves in quartiles for the four subgroups are shown in Figure 9. The designation A is for the first part of the experiment and B for the second. The letter I stands for the interpreted part and F for the formal part. Thus, IA refers to the data for the group which began on the interpreted part and IF to the second part of their experiment, when they were given the formal material. As is evident from the figure, the Vincent curves are relatively stationary over the first three quartiles, but all four subgroups show a definite tendency to increase in the last quartile before reaching criterion. I shall have some additional remarks to make about this experiment in a moment.

At this point, however, I want to summarize the kind of results obtained in the three experiments we have examined. The figures, as well as the related statistics, which I have not presented in quantitative detail here, strongly support the all-or-none model versus the simple incremental model. The primary reason for this is that in the three experiments we have examined, the Vincent learning curves are relatively stationary for the first 50-60% of the data prior to the last error. Additional experiments that qualitatively support these same conclusions have been recently analyzed in Suppes and Ginsberg (1963), and I do not want to repeat the ground covered there. Ginsberg and I did propose a simple two-stage model that accounts for the concave upward Vincent learning curves that we seem to be finding as characteristic of simple concept formation in children. The concavity is obtained by postulating a differential rate of conditioning for the two stages, in particular by postulating that conditioning in the first stage proceeds more slowly. When we move from a single-stage to a two-stage model, we are able to obtain a very satisfactory fit to the response data of these three experiments. Unfortunately, it is
not possible within the scope of this paper to present the quantitative details of this fit.

In discussing the pattern experiment performed in Brussels, I mentioned I would return to the response-latency curve shown in the top half of Figure 4. You will recall that there was an increase in latency in the last two sections before criterion was met. This is not an isolated finding. For instance, we have obtained the same sort of result in paired-associate experiments with college students, and Sheldon White of the University of Chicago has obtained a corresponding increase in eye movements just prior to criterion in concept experiments with young children. The qualitative features of these latency results are also well accounted for by a two-stage model. However, in propounding the virtues of a two-stage model for the experiments we have considered thus far, I do not mean to imply that I think a model or theory of this simplicity is adequate for the deeper problem of understanding the mechanisms by which children, or adults, form mathematical concepts.

**Some Results Concerning Transfer**

There are at least two senses in which we can talk about transfer. The first is in the sense of stimulus generalization of the sort required in order for a concept to be learned. Most of the experiments I am discussing in this paper require this sort of transfer as part of the training trials, because a different stimulus display is presented on every trial.

A second sense of transfer in concept-formation experiments is that of transfer from one concept to another. It is to be expected that positive and negative transfer between concepts occurs in the same sort of way that it occurs among stimuli. This second sense is the one I want to consider here. For this purpose, I turn to a fourth experiment, performed in collaboration with Rose Ginsberg. We have published some of the results elsewhere (Suppes & Ginsberg, 1963), but not the results to be mentioned here. The learning tasks involved in this experiment were equipollence of sets, that is, the concept of two sets having the same number of members, and the two related concepts of identity of sets and identity of ordered sets.

Ninety-six first-grade subjects were run in four groups of 24 each. In Group 1 the subjects were required to learn identity of sets for 56 trials and then equipollence for a further 56 trials. In Group 2 this order of presentation was reversed. In Group 3 the subjects learned first identity of ordered sets and then identity of sets. In Group 4 identity of sets preceded identity of ordered sets.

The sets pictured by the stimulus displays consisted of one, two, or three elements. On each trial two of these sets were displayed. The apparatus was as shown in the top half of Figure 10. Minimal instructions were given the subjects to press one of the two buttons (A₁) when the stimulus pairs presented were "the same" and the other button (A₂) when they were "not the same." When the correct response was made, the appropriate one of the two reinforcing lights, E₁ or E₂, flashed.

Our empirical aims in this experiment were several. First, we wanted to determine if the learning of simple set concepts by children of this age took place on an all-or-none conditioning basis. Second, we wanted to know if the learning of one kind of concept would facilitate or inhibit the learning of another, and if there were significant differences in the degree of transfer.

Third, we were concerned to consider the question of finding the behavioral level at which the concepts could be most sharply defined. For example, in learning the identity of sets, could the learning trials be most satisfactorily analyzed from the standpoint of all trials falling under a single concept; or would it be better to separate the trials on which identical sets were presented from those on which nonidentical sets were presented in order to analyze the data in terms of two concepts? Or
would a still finer fourfold division of concepts be
called for in terms of pairs of sets identical in terms
of order (O in the lower left part of Figure 10),
pairs of sets identical only as nonordered sets (10),
pairs of sets that are equipollent but not identical
(1E), and pairs of sets that are nonequipollent
(E)? At the lower right of Figure 10 is shown the
natural ordering of these concepts of identity, with
O, the strictest sense of identity, on the left and E
on the far right. The vertical line indicates the
point of division for the ordinary concept of identity
of unordered sets. O-type pairs and 10-type pairs
are identical in this sense. E-type pairs are not.

With reference to the earlier discussion of all-or­
one processes, I remark parenthetically that the
results for this experiment agreed very much
with those already discussed. Concerning transfer
results a fairly large number of detailed remarks
can be made about the data of the experiment. I
would like to restrict myself to one that I think is
of considerable significance. This concerns the evi­
dence for negative transfer.

The single case I want to consider is what hap­
pens when we train children on identity of ordered
sets before we train them on identity of sets in the
ordinary sense, i.e., on unordered sets. Using fairly
simple principles of interference confirmed in a
large variety of work on paired-associate learning,
we might expect negative transfer in going from
identity of ordered sets to identity of sets. O-type pairs and 10-type pairs
are identical in this sense. E-type pairs are not.

The line of analysis I have sketched for this ex­
periment represents a strong conviction of mine,
which I would like to express in a rather round­
about way. For anyone interested in the psycho­
logical foundations of mathematical concept forma­
tion it is natural to ask what is the sort of connec­
tion that holds between the logical structure of
mathematical concepts and the psychological pro­
cesses of acquisition of the concept. As far as I
know, not very much has been written in the psy­
chological literature about this kind of question.
My present view, based partly on our experiments
and partly on conjecture, is that the psychological
stratification of mathematical concepts will seldom,
if ever, do violence to the logical structure of
these concepts; but it will markedly deviate from the
mathematical analysis of the same concepts with
respect to the amount of detail that must be con­
sidered. Let me give a simple example. Even in
the detailed discussions of mathematical concepts
familiar among mathematical logicians, the physi­
cal size of the stimuli used in the language of math­
eematics to symbolize concepts, or the stimulus simi­
larity between the physical objects used as an
alphabet, is not discussed as a proper mathematical

Mathematical Concept Formation

Pairs of sets already defined, the 10 pairs and only
these require a different response in passing from
the concept of identity of ordered sets to that of the
identity of unordered sets. As may be seen in Fig­
ure 11, all the negative transfer is indeed isolated
in the pairs of the 10 subconcept. For the other
three subconcepts there is in fact slight evidence
of positive transfer, which accounts for the reduced
difference between the two overall curves on the left
as compared to the two for the subconcept on the right.

The single case I want to consider is what hap­
pens when we train children on identity of ordered
sets before we train them on identity of sets in the
ordinary sense, i.e., on unordered sets. Using fairly
simple principles of interference confirmed in a
large variety of work on paired-associate learning,
we might expect negative transfer in going from
identity of ordered sets to identity of sets. O-type pairs and 10-type pairs
are identical in this sense. E-type pairs are not.

The line of analysis I have sketched for this ex­
periment represents a strong conviction of mine,
which I would like to express in a rather round­
about way. For anyone interested in the psycho­
logical foundations of mathematical concept forma­
tion it is natural to ask what is the sort of connec­
tion that holds between the logical structure of
mathematical concepts and the psychological pro­
cesses of acquisition of the concept. As far as I
know, not very much has been written in the psy­
chological literature about this kind of question.
My present view, based partly on our experiments
and partly on conjecture, is that the psychological
stratification of mathematical concepts will seldom,
if ever, do violence to the logical structure of
these concepts; but it will markedly deviate from the
mathematical analysis of the same concepts with
respect to the amount of detail that must be con­
sidered. Let me give a simple example. Even in
the detailed discussions of mathematical concepts
familiar among mathematical logicians, the physi­
ical size of the stimuli used in the language of math­
eematics to symbolize concepts, or the stimulus simi­
larity between the physical objects used as an
alphabet, is not discussed as a proper mathematical

Mathematical Concept Formation

Pairs of sets already defined, the 10 pairs and only
these require a different response in passing from
the concept of identity of ordered sets to that of the
identity of unordered sets. As may be seen in Fig­
ure 11, all the negative transfer is indeed isolated
in the pairs of the 10 subconcept. For the other
three subconcepts there is in fact slight evidence
of positive transfer, which accounts for the reduced
difference between the two overall curves on the left
as compared to the two for the subconcept on the right.
part of the logical foundations of a given mathematical system. Yet for detailed investigations of learning, considerations of size and similarity of stimuli are of critical importance. In mathematical terminology, the mathematician is interested in much broader equivalence classes of stimuli to symbolize his working concepts than is the psychologist investigating the processes by which someone learns these mathematical concepts. Consider again for a moment the experiment on the identity of sets. I have tried to show the advantage of analyzing this experiment from the standpoint of four natural subconcepts, but even these four subconcepts do not exhaust the possibilities. There is a still finer substructure of 11 concepts generated by the four subconcepts and the number of elements in each pair of sets displayed. Unfortunately, a larger experimental group is needed to provide adequate data to pursue in detail the analysis of this finer substructure. A rather arduous but I think important experimental program we hope to carry out is the pursuit of these substructures for various types of mathematical concepts to the point where at the given level of training and development of the child, equivalence classes are formed such that no significant behavioral differences are detected by considering any finer substructure. I cannot claim that we have yet done this adequately for any of the mathematical concepts we have considered, but the potentialities of mathematical concepts for investigations of this kind are apparent when the usual psychological experiment on the identity of sets I have described is compared with the usual psychological experiment in concept formation, for which there is little possibility of defining any sort of substantial substructure.

As a final comment on this matter, it seems to me that behaviorally identified substructures of the sort natural for the identity-of-sets experiment are of interest not only for the psychological investigation of concept formation, but also for the logical foundations of mathematics itself. I do not have space to amplify this point in detail. The central point, however, is that at the present time many mathematicians concerned with the foundations of mathematics are unsatisfied with the basic concepts of logic and set theory that have been the focus of most foundational investigations since the turn of the century. A behaviorally oriented way of looking at mathematical concepts promises a new approach to the foundations of mathematics itself.

After this philosophical digression, I now turn to some more experimental data. As a natural example in which to expect positive transfer, let us consider again the third experiment, the one concerned with learning logic either in interpreted or formal presentation. Consideration of the mean trial of last error for each group on each part shows that there was positive transfer from one part of the experiment to the other for both groups. For example, the group that began with the formal material had a mean trial of last error of 14.1 on this part, but the group who received this material as the second part of their experiment had a smaller mean trial of last error of 10.9. In the case of the interpreted part, the group beginning with it had a mean trial of last error of 18.3, but the group that received this material after the formal part had a mean trial of last error of 7.7, a very considerable reduction. Now one way of measuring the amount of transfer from one concept or presentation of mathematical material to a second is to consider the average mean trial of last error for both concepts in the two possible orders. If we look at the logic experiment from this standpoint there is a significant difference between the group beginning with the formal material, completely uninterpreted as to meaning, and the group beginning with the interpreted material. The average trial of last error on both parts of the experiment for the group beginning on the formal part is 10.9 and for the group beginning on the interpreted part is 14.6. Moreover, these same comparative transfer effects are reflected in the average latencies in quartiles prior to criterion for both parts of the experiment and for both groups as shown in Figure 12.
part of the experiment for both the interpreted and formal group, that is, IA and FA, are shown on the left side of the figure and latencies in quartiles over the second part, IB and FB, on the right side. Average latencies in seconds are plotted on the ordinate. As in the case of mean trials of last error there is positive transfer from one part of the experiment to the other for both groups but, as is apparent from Figure 12, the average latency over the whole experiment is considerably less for the group beginning on the formal part than for the group beginning on the interpreted part. These results fly in the face of much pedagogical practice, even though they were suggested by our own experience in a limited pedagogical experiment. I have some theoretical ideas to account for these differential transfer results, but they are too tentative yet to be put on record.

GEOMETRIC INVARIANTS OF PERCEPTUAL SPACE

I turn now to the topic of geometric invariants and to a fifth experiment motivated by a classical program of research in geometry. At the end of the nineteenth century the famous German mathematician, Felix Klein, formulated his Erlanger program for characterizing the mathematical properties that remain invariant under a given group of transformations of a geometrical space. In the familiar Euclidean geometry, the properties studied remain invariant under the group of rigid motions, i.e., transformations of space which translate the origin, rotate the orientation of the frame of reference, and possibly reflect a coordinate system from a right-handed to a left-handed one, or vice versa. Topology, in contrast, is the study of the properties of objects that remain invariant under the much wider group of one-one bicontinuous transformations, of which the group of rigid motions is a very special case. Our aim in the experiment was to begin to investigate the geometric properties children consider most invariant. In this initial investigation we restricted ourselves to equilateral triangles, squares, and regular pentagons, and thus automatically restricted ourselves to plane geometry. In contradistinction to the other experiments I have described, no correction procedure of any sort was given. The child was simply asked to point to which of two triangles, for example, seemed most like a standard triangle. Each of the two triangles he was given to choose between was rotated or stretched a certain degree with respect to the standard figure. A typical stimulus display presented the child is shown in Figure 13. The standard stimulus shown at the top of the figure is an equilateral triangle with horizontal base, which may be inscribed in a circle whose radius is \( \frac{3}{4} \) inch. The triangle in the lower left has its base rotated 15° and it is stretched by a factor of 2.5. The triangle on the lower right has its base rotated 45° and its size stretched by a factor of 1.5. The task for the child was to indicate which one of the two lower triangles he judged most like the standard stimulus shown at the top. Each child was presented with two series of stimulus displays, run in two separate sessions. The two series together consisted of 60 triangles, 36 squares, and 18 pentagons. I present here only the data for the triangles, but the data for squares and pentagons are of precisely the same sort and do not deviate from the triangle data in any serious respect. The experiment was performed in collaboration with Marcia Bandy. We collected a large amount of data from children of widely varying ages and cultural backgrounds. I have space to mention only one or two of the most important aspects of the results. We first ran groups of first graders, fourth graders, and medium-ability and high-ability sixth graders in a typical California elementary school. From these four classes we obtained quite a striking developmental curve. There was a very strong tendency for the first graders to choose the triangle with minimum rotation independent of any other factors. This tendency uniformly and significantly decreased from the first grade to the fourth grade, from the fourth grade to the middle-sixth group, and from the
middle-sixth group to the high-sixth group. At the time we obtained these data, I had just been reading Stuart Sutherland's (1961) excellent survey monograph *Shape Discrimination by Animals*. The experiments he surveys from the literature clearly indicate that animals transfer very badly when the stimulus pattern is rotated and have far less difficulty when the same stimulus pattern is simply changed in size, which are just the sort of results we were obtaining from our first graders. It seemed that the base point of our developmental curve was tied in very nicely with a large body of animal data. But as is often the case with scientific experimentation, one is not content to let well enough alone. We also decided in conjunction with some other experiments we had been performing as part of a small pedagogical program in Ghana to see what some of the Ghanaian children would do in this same experimental situation, particularly classes coming from families with a high rate of illiteracy. The Ghanaian data shot our simple hypothesis about development into very small pieces, for the Ghanaian children of this age, who are just beginning instruction in English as a second language, do not make much progress in reading during their first year.

It is amusing to note that the standard geometry of the schools runs contrary to our results for these early grades. The Euclidean notion of congruence is invariant under rotations, but, of course, not under size changes. For this reason alone, the results of this experiment have already had small repercussions in some of our pedagogical work in elementary school mathematics.

**Mechanisms of Concept Formation**

From a mathematical standpoint, there are a number of equivalent ways of describing the sixth and final experiment I want to mention. As I indicated at the beginning, this final topic is concerned with the mechanisms of concept formation. The experimental setting in which we have been investigating these matters grew out of an attempt,
as yet unsuccessful, to study the learning of recursive properties. The idea of studying the learning of such properties was itself a natural generalization from conversations with a number of people on how children are able to learn something so complicated as a recursive grammar in learning to speak a first language. If we wish to stay within this framework, the experiment may be described in terms of how children learn a simple grammar for a finite number of expressions.

In any case, the stimulus material presented to subjects consisted of all strings of length three that may be made up from a two-letter alphabet (D and F). These strings are similar to the stimulus material in the proof experiments, with the 1's and 0's presented as red and green squares. There are, of course, eight strings of length three that can be made up from a two-letter alphabet. The child is told that the experimenter has a way of dividing the eight cards into two groups, and that he is to figure out the method by trying his own division, putting one group on the left and the other on the right. In the particular experiment to be reported on here the subject at this point was also shown two cards, correctly classified, as indicated at the top of Figure 15. In the bottom half of the figure I have shown a typical classification of the eight cards consistent with the information given on Trial 1 as shown at the top. The four-four split of the eight cards was a characteristic of most trials, although it was not encouraged in the experimental instructions nor required in order to be consistent with the information given the subject until the terminal trials of a given problem. On each trial the subject was shown the correct classification of another card, so every problem necessarily terminated at Trial 7 if it had not been solved on an earlier trial. For all problems the correct solution consisted of one of the six letter-specific hypotheses as we have termed them, i.e., the solutions requiring all cards that have a specific letter in the first, second, or third vertical position to be placed on the left, the remaining cards on the right. For instance, a problem might require that all cards that have the letter D in the second position be placed on the left and consequently that all cards that have the letter F in the second position be placed on the right. Each child was given problems of this type in 25-minute sessions on 2 consecutive days. The criterion for terminating a session, prior to the end of the second day, was that of solving four consecutive problems correctly on the first trial. (Considering the large number of responses that may be made in terms of classifying eight cards—there are 256 possibilities—this is a rather strict criterion of learning.)

The subjects participating in the experiment were 89 fourth-grade children and 46 sixth-grade children from elementary schools near Stanford. The experiment was performed in collaboration with Irene Rosenthal.

One further point about the design of the experiment. The two cards whose correct classifications were shown prior to Trial 1 on each problem were so selected that if the subject used on Trial 1 a letter-specific hypothesis of the sort just defined, then he solved the problem on that first trial. From the description I have given of the experiment, it should be evident how it was intended to bear upon study of the mechanisms of concept formation. The central idea is to elicit hypotheses behaviorally by the kind of response required, and
thus get one step closer to the concepts or hypotheses the subject is sampling, forming, or generating—whichever term you prefer—in reaching a solution of the problem.

A natural first question about the data is this: Do the hypotheses or concepts elicited as responses seem to be following as a first approximation the all-or-none model discussed earlier? The answer seems to be "yes." In Figure 16, backward learning curves from the trial of last error are shown for responses on Trial 1. Blocks of two problems are plotted on the abscissa and, as usual, the probability of a correct response, in this case a correct hypothesis, on the ordinate. The curves are for the fourth- and sixth-grade subjects who reached criterion. There is some tendency for both curves to decrease from left to right, but considering that the highest point reached is slightly less than .4, the evidence strongly favors an all-or-none over a simple incremental model. The curves lie between .2 and .4, and there is a large discontinuity between this range and the criterion of 1.0.

Additional supporting evidence for this conclusion is provided in Figure 17. In this case we have plotted the mean trial for solving each problem for the fourth graders who met criterion and also for all fourth graders. Problems are on the abscissa and mean trial of last error on the ordinate. I would not want to claim that either of the curves shown is horizontal, but their slopes are slight and, particularly in the case of the criterion subjects, support an all-or-none model as a first approximation.

On the other hand, subjects were not simply randomly sampling hypotheses from the total set consistent with the information available on each trial. Most pronounced was the tendency to learn early always to use a four-four split, as required for the correct solution. For example, for criterion sixth graders the mean problem for the last deviation from a four-four split was 3.3 and the mean problem of last error was 6.4.

I would like to close with a final comment on this experiment. In a related experiment, very similar in design to the present one, Madeleine Schlag-Rey and I found a pronounced hierarchy of hypothesis classes being used by college students. Moreover, some fairly simple operations in terms of intersection and union of sets can be defined on the most salient members of the hierarchy to lead in a natural way to the less salient but necessary-for-solution members. We tried this same sort of analysis on the present experiment, but with little success beyond identifying the obvious four-four-split property just discussed. The vast majority of the incorrect hypotheses used prior to solution of any of the problems was drawn from the set of hypotheses that call for such a split. But examination of the distribution of responses within this set is disconcertingly and surprisingly close to a uniform distribution. Other than this even split, there seem to be no highly salient properties strongly biasing responses prior to a correct solution.

This result leads me to the following speculation about concept formation. The search for a detailed algebra of concepts to provide explicit and definite mechanisms for generating new concepts out of old ones may in many situations, perhaps particularly with children, be a mistaken venture. With the exception of a few salient features, new concepts may be formed by random choices, and only after one or more instances of the concept have been reached or put together by accident or chance is the new concept recognized. To change Einstein's comment on quantum mechanics, "Perhaps after all, God is playing dice."

REFERENCES


SUPPES, P., & GINSBERG, ROSE. 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.