Reprinted from
PROCEEDINGS OF
THE 1964 INTERNATIONAL CONGRESS FOR LOGIC, METHODOLOGY AND PHILOSOPHY OF SCIENCE
Held in Jerusalem, August 26—September 2, 1964
Published by North-Holland Publishing Company, Amsterdam.

THE KINEMATICS AND DYNAMICS OF CONCEPT FORMATION*

PATRICK SUPPES
Stanford University, Stanford, California, U.S.A.

1. Introduction

Analyses of concept formation can be found in disciplines that seem superficially unrelated. The two oldest traditions are in philosophy and mathematics, the one reaching back to Plato and Aristotle and having a continuous history in the theory of knowledge, and the other going back to Eudoxus, Euclid and their several mathematical contemporaries and successors. The logical status of the analyses of concept formation given by philosophers ranging from Aristotle through Hume to Kant and on to Russell has a complex and ambiguous history. It is common in contemporary discussions, for example, to say that Hume was badly confused about the distinction between the logic and psychology of concept formation, but it is also characteristic of the people who say these things that they do not offer a very precise or literal definition of the logic of concepts, and the word "logic" is used by them in a way that is itself tantalizingly vague.

In the case of the logical analyses of concepts in a mathematical context, particularly as questions have come to be put in terms of precisely characterized notions of definability, a quite finished logical theory of concept formation has developed. Given any theory and given a concept it is possible to ask in a quite definite and precise way whether or not this concept is definable in the theory or, given the concepts of a theory, it is possible to ask if one of the concepts is definable in terms of the others. It is true that problems of definability have often not been discussed as problems of concept formation, and yet it is obvious that there is a close logical relation between the two subjects. If a concept is not definable in terms of a set of other concepts, then in one sense that concept cannot be formed from them. For example, by application of Padoa's classical method for establishing the independence of concepts it is easy to show that for most standard axiomatizations of classical particle mechanics, the concept of mass cannot be defined in terms of the concepts of particle and position, Mach's famous proposal to the contrary. The ordinary-language philosophers who talk about the logic of concepts

* This paper has grown out of research supported by the U.S. Office of Education, Department of Health, Education and Welfare.

are not talking about the application of methods like those of Padoa to the solution of well-defined problems, and it is my own suspicion that there exists no well-defined subject matter corresponding to much of their discussion of the logic of concepts, unless it is indeed the psychology of concept formation.

In spite of the temptation, I do not here attempt to make a case for the dissolution of the logic of concepts into the psychology of concept formation, but rather concentrate on a critique of the current status of concept formation in psychology, with particular reference to questions that seem to have philosophical interest.

I would like to end this paper with at least a sketch of a detailed scientific solution to the problems of concept formation along the lines laid down by Hume in Book I of his *Treatise*. Unfortunately one does not have to dig very far into the psychological literature to find that nothing like an adequate solution has been found. On the other hand, there is one aspect of the problem that I think is now quite well understood, namely, what I have termed the kinematics of concept formation, and in the next section I give a survey, albeit brief, of the main results now available.

2. Kinematics of Concept Formation

Within mechanics, kinematics refers to the descriptive theory of motions. Because of the remoteness of most teaching of mechanics to any complex problems of data analysis, it is often not realized that from an empirical standpoint kinematics can be a quite complicated subject. It was, for example, and still is no simple task to decide from astronomical observations what closed figure represents to a high degree of accuracy the orbit of any one of the planets. A fully detailed statistical discussion of the question is extremely sophisticated and certainly is never mentioned in any of the standard textbooks on mechanics. The corresponding descriptive theory of concept formation has received a great deal of analysis in the recent psychological literature on learning, and a quite reasonable account in descriptive terms of the learning of many concepts can be given. The intended meaning of "reasonable" and "descriptive" needs remarking upon before these terms can mean much to those unfamiliar with the recent psychological literature. In the first place, it is a characteristic of the recent learning literature to abandon the hope of giving a deterministic description of the process of forming a concept on the part of an organism and to settle for a probabilistic description, but it is a mistake to think that it becomes a simple matter to find an adequate probabilistic description as opposed to a deterministic one. In actual fact, for large bodies of data it is a demanding task to

satisfy with any strictness tests of goodness of fit. Moreover, for large bodies of data for which a probabilistic descriptive theory is postulated, many probabilistic relations assume a deterministic character at one remove from the data via application of the law of large numbers.

To make matters more concrete, it will perhaps be wise to sketch one simple experiment and the kind of descriptive theory applied to it. The experiment is one in which a young child is learning the concept of identity of sets. The children were of ages running from five to seven years. The sets depicted by the stimulus displays consisted of one, two or three elements. On each trial two of these sets were displayed. Minimal instructions were given the children to press one of two buttons when the stimulus pairs presented were "the same" and the alternative button when they were "not the same". In order to prohibit explaining the learning of the concept by a simple principle of stimulus association, a different stimulus display was shown on each trial. Because of this change of the stimulus display on each trial, no models at the level of simple stimulus associations can be applied to the response data of the children in any straightforward fashion. However, if we move from a stimulus-response association to a concept-association, the simple models used in quite elementary and primitive stimulus-response experiments work extremely well. Perhaps the simplest model is a so-called one-element model which postulates that the subject enters the experiment in the unconditioned state, i.e., the appropriate association or connection between the concept and the correct response is not established. On each trial, there is a constant probability c that the correct association will be established between the concept and the response, and thus that the subject, in this case the child, will enter the conditioned state. When the child is in the unconditioned state there is simply a guessing probability p of making a correct response, but when the conditioned state is entered, the probability of a correct response is one. A simple matrix may be used to describe transitions from the unconditioned (U) to the conditioned state (C).

	C	U
C	1	0
U	c	$1-c$

Other assumptions of a simple and natural sort are added to what has been stated in order to make the postulated sequence of conditioning states a first-order Markov chain. (It is worth noting, however, that we do not have such a chain in the observable responses themselves.) Once we are given the

guessing probability p and the conditioning probability c, then all probabilistic questions about the response data are uniquely and completely determined. This means that after estimating these two parameters from the data a wide variety of predictions maybe made.

The strongest prediction of the one-element model I have just described is that prior to the last response error, there is no evidence of learning. It is a characteristic of the model that the guessing probability p is constant prior to the last error. In contrast to this assumption, the central assumption of the simple linear incremental model is that there is an increase in the probability of a correct response on each trial. The simplest way to formulate this incremental model is the following. Let p_n be the probability of a correct response on trial n. Then the probability of an error, q_n, is simply $1—p_n$. It is postulated that $q_{n+1} = \alpha q_n$, where α, the learning parameter, is a real number between 0 and 1. A number of experiments on which these two models have been compared are described in Suppes and Ginsberg (1963). (For some related applications to concept identification see Bower and Trabasso (1964).)

Although the bulk of the simple experiments on concept formation favor very strongly the one-element model, there are several situations in which a compromise between the two most satisfactorily explains the observed response data. This compromise consists in postulating that instead of having simply a single element that is conditioned or unconditioned, the concept-response association is best represented by a two-element model. The two elements may be interpreted as aspects or characteristics of the concept itself. A number of different formulations of two-element models have been published in the literature; a typical and simple extension of the one-element model is that described by the following matrix:

	2	1	0
2	1	0	0
1	b	$1-b$	0
0	0	a	$1-a$

Here, the conditioning parameters a and b play the role of c in the one-element model. It is assumed that the subject starts in the unconditioned state with 0 elements' conditioned as represented in the matrix by the 0 state. The probability of moving from the state of 0 elements' being conditioned to the state of 1 element's being conditioned is a, and correspondingly the probability of moving from state 1 to state 2 is b. Moreover, the probability of a correct response when in state 0 is p_0, and the probability of giving the

correct response when in state one is p_1. As before, the probability is one of giving a correct response when all elements are conditioned, i.e., when the state is 2. It should be apparent that part of the greater success of this two-element model is simply the fact that it has four parameters, namely, a, b, p_0 and p_1, to be estimated from the data rather than two, as in the case of the one-element or linear incremental model. All the same, independent of parameter estimations there are some qualitative features of the data in many simple concept formation experiments that support the two-element model. For example, considerable evidence is presented in Suppes and Ginsberg (1963) to show that for many experiments the mean learning curve for response data is concave from above and quite apart from estimation of any parameters, such a curve is consistent with the two-element model, but not with the one-element or linear model. The essential point for the present discussion is that the one-element or two-element sort of model does predict with considerable accuracy the probabilistic characteristics of response data in simple concept formation experiments.

A typical prediction of the one-element model is shown in Table 1. These data are drawn from an experiment in which six- and seven-year old children were

TABLE 1.

Empirical and Theoretical Frequency Distribution of Response Errors in Blocks of Four Trials for Children's Learning of A System with Four Production Rules

Number of Errors	*Empirical Frequency*	*Theoretical Frequency*
0	9	8.15
1	59	58.30
2	161	156.46
3	172	186.63
4	92	83.48

being taught the simplest sort of mathematical proofs (Suppes (1961),(1964)). The experiment was performed in collaboration with John M. Vickers. The mathematical systems used in the experiment deal with production of finite strings of 1's and 0's. The single axiom is the single symbol 1. The rules of production are of the simplest sort. For example, given a string then one rule permits the addition of two 1's on the right. Another permits the deletion of a 1 on the right. The language used with the children was not, as you might

expect, that used here. 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 on the left in the lower panel was always illuminated red, corresponding to the single axiom, 1. Corresponding to each rule of production the child was given a button that he could use to light up additional squares or remove squares from the lower panel. His problem was to match the lower panel to the top panel. The theorem being proved was shown in the top panel. Each child was presented with 17 theorems per session for a total of 72 trials. The one-element model predicts that prior to learning how to use the rules of production the child simply guessed the correct response. Moreover, these guesses are drawn from the binomial distribution with parameter p. Table 1 compares the theoretical and empirical distributions for the number of errors in blocks of four trials, for one of the two groups in this experiment. The predictions are quantitatively quite good, and on a standard chi-square test the differences are, as you might expect, not significant.

I have chosen just this one sample of data. Many other similar instances can be found in the recent literature. The kind of predictions exemplified by Table 1 are the sort of descriptive predictions I have labeled *kinematical* in analogy with mechanics. What Hume attempted in Book I of his *Treatise* and what we all desire, namely, an adequate causal explanation, is certainly not given by the kinematical theory of the one-element and two-element models I have described thus far. I now turn to this more complicated problem.

3. Dynamics of Concept Formation

From a philosophical standpoint, the solution to the "kinematical" problems of concept formation are of only limited interest, just as in the case of mechanics it is dynamics and not kinematics that has stimulated so much philosophical discussion of the nature of scientific theories in physics. In many respects, a paradigm example of a dynamical theory of concept formation is provided by Hume in Section VII of Book I of his *Treatise*, the section treating abstract ideas. Following Berkeley, Hume attempts to reduce the formation of abstract concepts or ideas to the process of collecting around a term a number of particular ideas. As one sort of modern discussion of these topics would put it, Hume was concerned to characterize the process by which abstract ideas are coded. To give a complete account of the coding process is certainly in one sense to provide an adequate dynamical or causal theory of concept formation. I said that Hume's theory is a paradigm example, but this is true only in broad outline. It is far from being a paradigm example in its lack of detail and the difficulty of developing

a substantial systematic theory from the general notions thrown out by Hume.

The modern theory closest to Hume is that which suggests that the process central to concept formation is the process of verbal mediation. There is a very extensive literature in psychology on verbal mediation, but if one scrutinizes this literature for the hard-core theoretical assumptions, it is difficult to find anything substantial that goes much beyond what Hume had to say. A good way of pin-pointing the problem of verbal mediation theories is to move on to the approaches that have arisen in attempts to solve various practical and theoretical problems involved in constructing intelligent machines. This approach to concept formation can probably most aptly be labeled the theory of artificial intelligence. The superficiality of our understanding of how concepts are formed immediately becomes evident when we examine what help any particular theory in concept formation can give us in programming a computer to play a reasonably adequate game of chess, or to solve simple perceptual problems of pattern recognition. The fact is that verbal mediation theory and its kin are too fuzzy and indefinite to provide any serious scientific help in solving these problems. We all can agree in general terms that there must be a coding process which the brain uses to represent concepts and to store information, but the details of how this coding process works have not been successfully elucidated in current theories of verbal mediation. It could of course be that this elucidation has taken place and the difficulty facing the scientist who wants to apply the theory to problems of artificial intelligence is that the computer he has at hand is not of adequate capacity, but this is not at all the situation. It is simply that the psychological theories of verbal mediation are lacking in systematic scientific content.

From a mathematical standpoint undoubtedly the simplest and neatest dynamical theory of concept formation would be one formulated in terms of an algebra of concepts. The intuitive idea is that an organism is able to apply certain operations to his repertoire of concepts at a given instant in order to produce a new concept. From a formal standpoint such a set-up would be characterized in terms of an algebra in which the elements were the initial concepts and the operations corresponded to operations the organism could perform. A natural first start is to think in terms of Boolean operations on concepts, but it does not take much additional reflection to make clear that this is certainly not an adequately rich apparatus for forming concepts of any complexity. The difficulty, of course, is evident at once when we consider what range of concepts can be defined by use of Boolean operations. Certainly we cannot build the imposing structure of concepts possessed by all higher organisms. The proof, if one is desired, follows by direct application

of Padoa's method, and indicates the kind of link that may be forged, once a systematic theory of concept formation is considered, between mathematical and psychological theories of concept formation.

One can continue to push the algebra of concepts by introducing a richer set of operations. There is unfortunately very little, if any, constructive literature to be cited on this line of development, but there is one line of attack that seems to be so intuitively promising that I want to describe it even if it is not clear at the present time how the details are to be worked out. I have in mind the single primitive binary relation of set theory, namely membership, and the operation on concepts corresponding to the membership relation. We know that from a mathematical standpoint it would be a very powerful method of attack. It is also clear that this approach has close connections with verbal mediation theory. The forming of a set, or the assertion that an object is a member of a set, corresponds closely in a psychological sense to the notion of establishing usage for a general term. Admittedly talk of sets of sets of sets does not have any clear psychological meaning or reference, but if we talk about a chaining of verbal mediators, as would arise from the successive notation for sets of sets of sets, we then have immediately at hand in the notation a device that can be linked to the theory of verbal mediation, and also to general ideas of coding. It is my own hunch that this is probably one of the most promising directions in which to work in developing an adequate dynamical theory of concept formation. On the other hand, many treacherous and difficult problems have got to be solved and it is certainly not clear at the present time how to solve them. One interesting aspect of this approach is that if it could be worked out adequately, I am sure it would have repercussions on the foundations of mathematics itself. From a psychological standpoint, talk about sets and algebraic operations sounds rather like medieval talk about mechanics. Not that the talk is wrong. It is just that it seems hopeless in this vein ever to achieve an adequate solution to the problem being investigated. In every case, psychologically we want to turn at once from "abstract" talk about a set to immediate and direct questions about how notation for these sets is coded, but the implications of this line of thought for the foundations of mathematics cannot be explored in the present paper.

As still another inadequately worked-out theory of concept formation, I would like to mention some recent work I have been pursuing with some younger colleagues (particulary Madeleine Schlag-Rey). The central ideal is to extend the kinematical models discussed earlier by imposing several levels of conditioning, the most obvious way of describing two levels being that of rule and instance conditioning. Let me illustrate this distinction by a simple example. Suppose a subject is asked to classify objects that exemplify

a number of complex properties, for example, shape, size, color and orientation. A simple example of a rule at one level would be the rule that the correct classification depends on exactly one of these properties. The instances in this case would be the various hook-ups between the positive and negative instances of each property and the classification. A second simple example of a rule, or as we sometimes say, second-order hypothesis, would be the hypothesis that exactly two properties of the list given above are required for correct classification of the objects. According to the theory we have attempted to apply to experimental data, it is postulated that conditioning of rules changes very slowly in comparison to the conditioning of instances and generally there is a high probability that most, if not all, of the instances of a given rule will be run through before the rule is rejected. An *a priori* probability distribution, to be used in the selection of a rule, is also postulated, and in fact the present evidence strongly points toward the desirability of assuming, and then attempting to work out the details of, a hierarchy of rules that is imposed by the organism on the basis probably of both past experience and innate abilities, in order to avoid combinatorial chaos—for example, the number of rules for a two-way classification of 100 stimulus items is just the number of subsets, i.e., 2^{100}, and no unstructured or brute-force attack on this number of rules is the least bit feasible.

From many standpoints the current central problem of concept formation is to find the principles that lead organisms out of the combinatorial jungle that is uncovered in any purely logical or mathematical analysis of complex problem solving. An understanding of these principles would lead to an enormous gain in our understanding of human thinking. And the present problems are not dependent for their solution on tomorrow's news from the neurophysiological front. It would, for example, be a big step forward to be able to lay down general principles for getting about with a computer in this combinatorial jungle of logical possibilities, even if the principles used were not at all those used by any living organisms. What we seem to lack are the right conceptual ways of looking at either concept formation or complex problem solving, and the finding of new and more powerful approaches is bound to have repercussions in philosophy because of the closeness of the subject matter to much of the classical tradition in the theory of knowledge, and the fundamental importance of the processes of concept formation for all human thinking and action.

REFERENCES

[1] BOWER, G. H., and T. R. TRABASSO, Concept identification. In R. C. Atkinson (Ed.), *Studies in Mathematical Psychology*. Stanford: Stanford University Press, 1964, pp. 32–94.

[2] SUPPES, P., Towards a Behavioral Foundation of Mathematical Proofs. Technical Report No. 44, Psychology Series, Institute of Mathematical Studies in the Social Sciences, Stanford University, 1961.

[3] SUPPES, P., Mathematical Concept Formation in Children. Technical Report No. 64, Psychology Series, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1964.

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