FORM DISCRIMINATION in YOUNG CHILDREN and the CONCEPT of SIMILARITY

John A. Taylor

Ph.D. Thesis, University of Edinburgh

September 1972

BIBL. UNIV.

I hereby declare that this thesis has been entirely composed by myself alone, and that no part of it has previously been submitted for a higher degree.

Acknowledgements

This work was begun while the author held an SSRC Research Studentship in the Cognition Project at the University of Edinburgh under the direction of Dr. Margaret Donaldson and Dr. Roger Wales. Their guidance and support have been invaluable. George Balfour, Robin Campbell, Robert Grieve and Valerie Horne have all, in their different ways, provided stimulating comment and criticism. Thanks are due also to the various head teachers who allowed their pupils to be investigated.

CONTENTS

SUMMARY

A brief account is given of differentiation theory as it relates to form discrimination studies, and relevant experimental work is reviewed. Some preliminary comments on the concept of similarity, arising out of this and pertaining to the extent to which children can be said to possess an adequate grasp of what is implied by "the same as," are made as an introduction to the first experiment. This examines the performance of a group of children twice, at mean ages 3 - 8 and 4 - 9, in matching-from-sample discrimination tasks with stimulus material varying in orientation only, and in form and orientation. Considerable improvement in terms of increased number of correct responses is shown. A sequence of three stages is described. In the first, characterised by a small number of correct responses and a small number of multiple responses (MR), that is responses where more than one comparison figure are matched to the standard, performance is affected as much by extraneous factors such as position of a figure in the comparison array as by stimulus features. The second stage shows an increase in both MR and number of correct matches; here global features of similarity appear to be being detected and used. The final stage, with most responses being correct and few MRs being given, represents the most competent level. It is indicated that failure to detect stimulus features, rather than the presence of a deviant notion of the meaning of "same,"/

"same," is responsible for the error patterns shown.

The concept of similarity is then examined in greater detail. A number of distinctions are drawn, in particular, between the specification of similarity relations between members of a stimulus set in terms of its attribute structure and the perceived similarity of the same set as expressed by subjects' judgements. The importance of providing a normative model as a baseline for the assessment of performance is emphasised. A number of ways of specifying stimuli are described, together with methods of analysing similarities data. The use of a model to link perceived and physical structure, and thus to give some indication of the processes underlying discrimination performance, is considered. A set of experiments is then described which embody these ideas, and it is shown that even in four-year-old children errors in a matching-from-sample task systematically reflect features of the whole stimulus set. It is also shown that this does not hold with pair comparison presentation. The capacity of children for redefining the attribute structure of stimulus sets is brought out. It is concluded that much more attention should be given to stimulus specification in formal terms if the processes involved in form discrimination are to be elucidated, as opposed to the discriminability of a particular set of stimuli under particular circumstances.

INTRODUCTION

This dissertation is organised around two major issues. The first concerns the experimentally-defined ability of children around the age of five to discriminate visually between two-dimensional forms varying in several attributes, and between forms varying only in the orientation in which they are presented. A learning paradigm has not been used, as the context in which this work was developed emphasised the study of the use by the child of the cognitive apparatus already available to him in extracting and processing stimulus information rather than the modification of responses over time following differential reinforcement. The major portion of the results to be reported consists of the confusion errors made in a matching-to-sample task and their distribution over both groups and individuals, in an attempt to provide at least some counterweight to the stricture that "psychology has not gone far enough in investigating the growth of ability to detect regularity, order and structure." (E.J. Gibson, 1969). The second issue is centred on the formal aspects involved in this kind of experiment; how best are the relations among the elements of the stimulus set to be defined in terms of their physical parameters, to what extent is the structure of the stimulus set reflected in the error patterns obtained and how is this to be assessed, and what sort of consideration should be reflected in the approach taken to each of these problems. To the extent that the second issue can be dealt with adequately, we should be able by bringing both lines together to make some general and principled statements about the child's perceptual functioning.

It/

It may be as well to point out that "perceptual" here is being used in the sense made explicit by Imai and Garner (1968), where they refer to the use of "... the direct perceptual properties of the stimuli ..." Examples of such properties are size, number of angles, or colour. They are contrasted with derived or learned properties such as those which would define certain geometrical forms as letters of the alphabet. Thus in talking of perceptual functioning we are concerned with the use made of immediately available aspects of the stimuli in matching them to each other.

The order of presentation of these topics is chronological rather than systematic, In that the formal issues are tested fully only in the second series of experiments to be described. The first two chapters, dealing with the earlier work, contain an account of the theoretical position adopted as a frame of reference in posing the initial questions, a brief review of the relevant literature, and an account of the initial experimental work. This was carried out as part of a much larger-scale study of the development of cognitive and linguistic skills in pre-school children. It was a collaborative venture as far as the initial choice of material was concerned, but the collection and analysis of results, and the implications drawn from them are the responsibility of the present author. The findings have already been published (Taylor and Wales, 1970) and the paper is included here as an appendix. There seems no point in giving basic data twice over, and therefore a condensed version is given in chapter two, together with a statement of the substantive issues raised.

Formal/

Formal issues are introduced in Chapter 3. These include a discussion of different approaches to stimulus specification, the use of non-metric multidimensional scaling methods in the analysis of similarities data, the nature and inter-relation of physical and psychological similarity, and the question of whether the study of individual or group results is more appropriate. A number of cross sectional studies, as opposed to the longitudinal one of chapter two, are then described, which are intended to illuminate the foregoing theoretical questions. In addition, the essentially non-random nature of the subjects' responses and the influence of the context provided by the total available stimulus set are brought out. Finally, and inevitably, the implications of the various aspects of the whole study are considered in Chapter 5.

Chapter 1

Preliminary Issues

Since "perception", in any theoretical or empirical context, refers to the interaction of an individual with features of his environment registered via the various sensory modalities, it is strange that an account of the topic firmly based on the closely and convincingly argued premise that structured, non-random and consistent information is available from the physical world about the objects and events which go to make it up did not appear prior to 1966 (J.J. Gibson, 1966). In his theory of information pick-up, it is assumed that the invariants of stimulus structure are registered by a perceptual system, rather than constructed by it. The developmental implications of this thesis have been elaborated by E. J. Gibson (1969). Progressive differentiation of features of stimulation, taken as "... increased specificity of correspondence between stimulus information provided by the environment and the organism's perception of it ..." is seen to be the modification which occurs in the course of development and learning. This theoretical approach is contrasted with what are described as enrichment theories; the common factor in these is their emphasis on some addition to sensory input in accounting for changes in perception. On the Gibsons' view the criterion for learning is increased specificity, in the sense that properties and structural features of stimulation which were not previously detected come to be responded to. Objects are differentiated in terms of the distinctive features which define the/

the unique characteristics of each in terms of its relation to other objects; a close analogy is drawn between this notion and the distinctive features of phonemes as described by Jakobson and Halle, whereby speech sounds are characterised uniquely in terms of the features which they do not share with the other members of the set. There is a clear emphasis on the relational nature of distinctive features; they are only relevant in a particular context. The processes underlying the refinement of differentiation are taken to be abstraction, of those aspects of stimulus information most apt for the task in hand; filtering, apparent in the increasing ability to ignore irrelevant stimulus. attributes; and attention, the orienting of receptor surfaces towards stimulus objects. The general prediction based on this set of ideas with which we shall be concerned here is that the fewer features there are differentiating between the members of a stimulus set, the greater will be the number of confusions between them.

The attraction of a theoretical basis which emphasises active search for and extraction of relevant stimulus information should be obvious for a study where considerations of prior knowledge of the task, the acquisition of appropriate learning sets, the presence of dimensional transfer or the nature of mediating responses are irrelevant. At the same time it should be noted that relatively little is said about the processes to which the information is subject once it has been extracted; here perhaps E. J. Gibson swings too far away from the schema theorists, exemplified in her eyes by Bartlett, Vernon, Piaget and Bruner, who concentrate on the cognitive, inferential aspects of perception without much consideration/

consideration of what activates the schema. However, it is possible to make more specific predictions on the basis of differentiation theory, and to that extent it can be considered to be a more useful basis for investigation; and although it emphasises the detection of stimulus information, the use to which this is put is not entirely overlooked.

With this brief exposition of the framework within which the results to be reported here are to be evaluated, it remains to review the experimental findings which influenced the design of the first study and to present a preliminary discussion of the concept of similarity as it enters into work of this nature. No claims for completeness are made, for either of these topics; the material quoted has been chosen simply to exemplify points of major relevance. Thus the literature on discrimination learning has been virtually ignored except where it meets this criterion.

Most of the work that has been done on form discrimination in young children has been concerned more with their ability to differentiate between identical forms in varying orientations than between distinct forms. Definite and consistent orientation preferences for simple geometric figures in four- to five-year-old children were shown by Ghent (1961), and Ghent and Bernstein (1961) demonstrated that this affected accuracy of response in a matching task. The figures in the former situation judged by the younger children to be "right way up" all had their "focal point", that part of the figure which catches the attention most readily, such as the point of a vee, at the top. This fact was interpreted in terms of a directional scanning hypothesis, whereby younger children/

children are held to fixate the focal point and scan downwards from there, whereas older children start scanning from the top, regardless of the position of the focal point. Braine (1965) (as Ghent had by this time become) tested this notion using geometric figures with focal points at one end and distinguishing features at the other. Her prediction, duly confirmed, was that younger children (mean age three years six months) would show improved matching to sample from a four-choice array with figures with the focal point at the top, and that older children (mean age four years nine months) would perform better with the distinguishing feature at the top. A third group of subjects, mean age four years two months, showed no comparable difference.

Ghent's version of the scanning hypothesis is contradicted by the findings of Kerpelman and Pollack (1964). Using irregular pentagons in a matching-from-sample task they concluded that differences in the bottoms of the forms provided most information for discrimination with four and five year old children. It is unfortunate that neither of these investigations included some record of eye movements; the work of Russian investigators (Zaporozhets, 1964) has shown that systematic scanning of the outlines of even simple forms does not appear before the age of five. Without a record of eye movements the scanning hypothesis must remain not proven.

The scope of enquiry is extended beyond this seeming cul-de-sac by a study by Wohlwill and Wienner (1966). They used figures differing in presence or absence of internal detail, being open or closed, and of high or low directionality, in a two choice matching task with orientation only being varied, with subjects of mean age four years two months. Left-right reversals (i.e. b-d) produced/

produced significantly more errors than up-down reversals (i.e. b - p) although mean error scores were rather low (5.4 out of a possible 32 per S). Response latencies were higher for the low-directionality figures than for the high, which suggests that problems involving the former require a greater degree of effort, in some sense, for their solution.

Working with somewhat older children (first and third grade, American) Rosenblith (1965) found that in a one-choice matching situation, the one comparison figure being changed until a match to the standard was reported, up-down reversal resulted in fewer confusions than left-right reversals. The figures used were right isosceles triangles and circles with 20° gaps in the circumference. Sekuler and Rosenblith (1964) showed that type of error in this same discrimination task is affected by the relative position of the stimulus figures; no errors are made in discriminating between pairs with the hypotenuse horizontal when they are presented side by side, but confusions do occur when the hypotenuse is vertical or oblique. Presentation of these figures one above the other produces the opposite error pattern. Similar results were obtained with the incomplete circles.

A discrimination learning study of some interest was carried out by Rudel and Teuber (1963) using a two-choice training situation with groups ranging from three years six months to eight years five months in mean age. Only one member of the youngest group failed to reach criterion with a vertical vs. horizontal line problem, and with problem 1 C, Fig. 1. Only three members of the two youngest groups succeeded in a left-right reversal problem (1 A, Fig. 1); and only one subject from these groups reached criterion in discriminating 45° oblique lines. Performance on all problems improved with age, but there were noticeable/

noticeable differences between the mean number of trials necessary to achieve success on the first two and last two problems described above at all age levels.

This paper provoked two studies directed against the authors' argument to the effect that because the human four-year-old performs only at the same level as an octopus in orientation discrimination, some part of the visual system of the brain must still not have reached full development by this age. Jeffrey (1966) by using a phased training procedure directed at the discrimination of oblique lines, found that half of his experimental group became able to differentiate between them, whereas none of his control group did. Mean age here was four years five months. Failure in discrimination cannot therefore be attributed solely to neurological deficit. Over and Over (1967) make the same point in a study comparing discrimination learning and matching procedures. With the former, results comparable to those of Rudel and Teuber were obtained; children of mean age four years one month had little difficulty with the vertical/horizontal line discrimination, but were much less successful with obliques. Similar results were found on the former problem using a two-choice matching technique, but performance with obliques was markedly improved. Presentation of the standard either above or beside the comparison figures did not affect the results. A second experiment with comparison figures exposed at varying intervals following presentation and removal of the standard led to the suggestion that poor memory for orientation differences is responsible for the worse performance found in discrimination learning tasks (cf. also Bryant, 1969).

Further evidence on the influence of stimulus alignment is provided in this context by Huttenlocher (1967). She repeated part of the Rudel and Teuber experiment,/

experiment, presenting up-down and left-right reversed figures one above the other as well as side by side. Two groups of five year olds were tested on the four sets of figures shown in 1 and 11, Fig. 1. All subjects solved the up-down problems in 1, and the left-right problems in 11; seven out of sixteen solved the left-right problems in 1, and thirteen out of sixteen the up-down problems in 11. Significantly more trials were required for A and B than for C and D in 1. Thus the relative position of such figures is shown to have a considerable influence on the results, but even taking this into account discrimination of left-right reversals is more difficult than of comparable up-down ones. Sekuler and Houlihan (1968) report a similar finding with adult subjects, the measure here being time to respond "same" or "different" correctly to pairs of tachistoscopically exposed figures.

The ability to discriminate between different forms appears to develop in advance of that required for discrimination of orientation differences alone. Birch and Lefford (1967) found the mean correct score on the discrimination sub-test of the 1937 edition of the Stanford-Binet to be 10.9 out of a possible 12 with five year old subjects, the most common error being failure to distinguish between upright and inverted triangles. Rosenblith (1965) points out that she used an orientation rather than a form discrimination task because of the failure of subjects in the age range of interest to make errors on the latter. Using random polygons of four, eight, twelve and sixteen sides in a series of 80 three-choice oddity problems, Brown and Goldstein (1967) found very few errors made by five year old children. One must admire the stamina of both the experimenter and the subjects here, as each session lasted some two hours. Response latency was used as a measure; a curvilinear relationship was found between this and number/

number of sides, with time to respond being shortest with eight-sided figures. It is suggested that this relation represents an interaction between the discriminability and the information content of the stimulus figures; simpler figures, being less free to vary than more complex ones, will be more difficult to discriminate. With increasing complexity discrimination becomes easier as figures become more variable until, beyond some optimal level represented here by eight-sided figures, more information than can be immediately processed is available. The hypothesis is a reasonable post hoc account of their findings, but the experiment adds relatively little to our knowledge of children's discriminative abilities.

The studies reviewed so far have been on a relatively small scale, with the possible exception of the last-mentioned, restricted in terms of both type of stimulus material and the number of comparison figures available to the subject. Particularly in the light of Garner's (1966) emphasis on the importance of the properties of entire stimulus sets in determining responses to problems involving only one member of the set, this raises the question of the extent to which the phenomena discovered can be taken as fully representative of discrimination performance. Two large-scale studies by E. J. Gibson and her colleagues do much to redress the balance (Gibson, E., Gibson, J., Pick and Osser, 1962: Gibson, E., Osser, Schiff, and Smith, 1964). The former involved a matching-to-sample test, with a comparison array of thirteen figures, carried out with subjects between four and nine years of age. It was devised as part of an investigation of reading skills. The standard figures were letter-like forms, and the twelve non-identical members of the comparison set represented systematic transformations/

transformations of the standard; the point of the experiment was "... to trace the development of letter differentiation as it is related to those features which are critical for the task ..." (E. Gibson, 1965). Of the twelve transformation types used, five altered the shape of the standard forms; three of these involved line-to-curve changes, or vice versa, with one, two or three lines being varied, and two were described as topological, either introducing a break in a closed standard or closing a gap already present in it. There were three rotation transformations and two reversal; with symmetrical figures the effects of the latter would be identical to those of certain rotations. The list was completed by two perspective changes, slant left and slant back, which produced minimal differences between the standard and the transformed figure. Subjects were instructed to pick from the array those figures which were "exactly like" the standard. Topological transformations presented little difficulty, even to the youngest subjects; confusions with perspective transformations, however, were numerous for all groups and showed little tendency to decline as age increased. Line to curve and rotation and reversal transformation produced error scores between these two extremes in the youngest group, but they declined sharply with age. In the first set at all ages scores reflected the number of lines altered, with changes in only one representing the greatest level of difficulty. A similar rate of decrease was found for rotations and reversals; 45° and 180° rotations were more likely to result in errors than 90° rotations, while the two types of reversal, left-right and up-down, were similar in the number of confusions they produced.

This/

This experiment was replicated, on the same group of five year old subjects, with letters of the alphabet as standards and the same transformations applied to them. Confusion errors correlated 0.86 over the two situations. The authors suggested, on the basis of their findings, that the influence of transformation is more important in predicting discrimination performance than are the characteristics of the standard. Improvement in discrimination with age comes from the learning of "... features or dimensions of difference which are criterial for differentiating letters." (Gibson et al 1962, p. 904); hence the relative failure to detect the non-criterial perspective changes, and the success with topological transformations. Thus break and close in figure outlines can be understood to act as distinctive features for all age-groups in this study; line to curve and rotation or reversal transformations are initially used much less as such, but become more salient as the child grows older and the probability of detecting the differences marked by them increases; and perspective transformations by and large have no differentiating value.

The importance of transformation type in predicting discrimination performance is demonstrated in an experiment by Pick (1965). She contrasted two hypotheses relating to the learning of discrimination; one that improvement depends on learning how the forms differ, the other that learning a schematic prototype of the standard is necessary. Using part of the stimulus material designed by Gibson et al (1962), she first taught kindergarten children to differentiate between the standards and their transformations. Three transfer tasks were used. In the first, corresponding to the first hypothesis, the subjects were tested on ability to discriminate between different standard forms from those/

those used in the training period and their transformations, which were the same as those to which they had already been exposed. The second transfer group were tested with the same standards but different transformations, and the third, control group were given different standards and different transformations. This was the order of success in the transfer test; both experimental groups showed significantly better discrimination than the control group, but the one tested with the same transformations was significantly better than the other. Thus the superiority of emphasising distinctive features rather than working from a template of the standard figure is demonstrated.

The second major study by Gibson and her colleagues (Gibson et al, 1964) was a direct attempt to discover what the distinctive features of letters are: "What dimensions of difference must a child learn to discriminate for each letter to become unique?" (Gibson et al, 1964, p.2). The method used represents an attempt to extend Jakobson and Halle's analysis of phonemic distinctive features to graphemes, or pictorial symbols. In the original, each phoneme is defined uniquely in terms of its feature pattern, that is by the presence or absence of each feature; to remain distinct over a wide range of speakers, etc. the features must be invariant under certain transformations and so have to be essentially relational. Two such feature lists were drawn up for this study, each of which provided unique descriptions for each of the letters of the alphabet. These were tested against the confusion errors made by four year old children in a matching-to-sample task with letters. The first experiment produced too few errors for comparison with feature list predictions to be possible.

The/

The second used both errors and response latencies with two sets of letter stimuli, of high and low confusability as defined by the feature lists. Under these conditions it was found that both errors and latencies were significantly higher for the high-confusion material. An admittedly crude comparison between number of confusions and feature differences produced a greater-than-chance number of significant correlations, thus supporting the notion of using a feature list in such a context. A multidimensional scaling analysis was carried out on the confusion matrix. This yielded three dimensions, straight vs. curve, relative obliqueness, and one which "... might conceivably be thought of as complexity." (Gibson et al, 1964, p.18). In addition open and closed letters were clearly separated in the final configuration.

There are relatively few firm conclusions which can be drawn from these studies. Ghent's work shows that orientation preferences exist and that they can affect performance in a matching task; her directional scanning hypothesis, however, is directly contradicted by Kerpelman and Pollack's findings. In spite of the amount of work done, it seems that this particular case presents a blind alley. The degree to which direction is marked appears to affect time taken to make a discrimination, as indicated by Wohlwill and Wiener's demonstration that response latencies are higher for low - than for high-directionality figures. They also show that left-right reversals are more likely to produce confusions than are up-down, as does Rosenblith. With two-choice arrays this error pattern may be reversed, however, with appropriate modification of the lay-out of the stimulus array. Discrimination between pairs of/

of oblique lines is well-nigh impossible for children under five in a learning task without some form of pre-training, or unless a matching situation is used. The results obtained by Gibson et al (1962), which failed to show any difference in difficulty between left-right and up-down reversals, are in contrast to the conclusion that the former provide a more difficult task. This apparent contradiction has been resolved by a reanalysis of the relevant parts of the data by Rosenblith (1965). She points out that only a few of the standard figures used allowed both types of transformation to result in differences from the standard, the remainder being symmetric around either the horizontal or the vertical axis, and that on these there were more confusions with the left-right transformation. Where rotation alone is concerned, 45° and 180° transformations are more confusing than 90°. Little can be said about shape discrimination, beyond the fact that it develops more rapidly than the ability to discriminate orientation. Features which are likely to be of use in this situation are open vs. closed, straight vs. curved, and presence vs. absence of oblique segments.

Turning to the question of similarity, there are two guises in which it is raised in this context. One concerns the theoretical and methodological issues involved in studying the concept of similarity as such and the way in which it is used in a given situation. Discussion of these problems is deferred until the results of the first experiment have been presented. The other is related to the practical question of how well children around five years of age understand and use the relation expressed in such terms as "the same as" or "just like", which/

which are of necessity used in the instructions in matching studies of the type just reviewed. Even for adults the notion of similarity is not unequivocal; indeed, Shepherd (1964) points out that a common response to the request to indicate which of several stimuli varying in more than one attribute is most similar to a standard stimulus is "Similar with respect to what?" There is evidence (Griffiths, Shantz, and Sigel, 1967) that five year old children do not use "same" correctly, and that this causes more difficulty than other relational terms such as "more" and "less". It is therefore extremely important to have as clear and unambiguous an indication as possible of the adequacy of children's ability to handle the notion of similarity if deductions about their perceptual abilities are going to be made on the basis of data gathered in situations involving the application of "same".

There are two ways in which errors might arise from the child's use of similarity in such a task. The "meaning of the word" may not be correctly understood, inasmuch as its reference to at least partial identity between two (or more) objects may not be appreciated. If this is so, then one would expect reaponses to be essentially random, or to be based on some such (experimentally) irrelevant factor as position in the array. In the other case the word "same" would be understood in its relational aspects, but its application might deviate from adult criteria. Here systematic error patterns would be expected, although different in structure from that which an adult might produce with comparable material. Such an independent assessment is available for the subjects of Experiment 1. They were also tested, roughly simultaneously with the discrimination/

discrimination task, in a classification task as part of which they were required to indicate objects which were "the same in some way" as one presented by the experimenter. Two types of material were used, common objects (e.g. toothbrush, egg-cup) and geometrical shapes. Two sets of each type were used. In one there were four kinds of objects and four colours, with the attributes of shape and colour being correlated; that is, all squares were red, all circles blue, etc. In the other the attributes were orthogonal, that is, there were three examples of each shape, each of a different colour. The experiment is described by Campbell and Young (1968); its linguistic aspects are examined by Donaldson and Wales (1968). It is quite clear from the results that, with two exceptions, the children were able to make appropriate choices on the basis of these instructions. Thus we can state that any errors in the discrimination task are not due to lack of comprehension of its relational aspects, although the question of deviation from acceptable adult standards still remains and will be taken up in the light of the results. The two exceptions will also be discussed in the context of their discrimination performance.

Chapter 2

Experiment 1

A developmental study of form discrimination in pre-school children

The object of this study was to examine changes over time in pre-school children's ability to discriminate visually-presented material differing in both shape and orientation and in orientation alone. The focus of interest is in the use made at different ages of such differentiating attributes of the stimulus material as are available, whether there is any systematic basis to the types of confusions made, and the extent to which response patterns can provide information on the interaction between the child's use of similarity relations and his observed perceptual functioning. At this stage experimental strategy was Baconian rather than strictly hypothetico-deductive; previous studies have been used to provide a basis for informal examination of a particular problem situation rather than as a source of testable hypotheses.

Method

For three reasons it was decided that a matching-from-sample technique, using fairly large arrays, should be used in preference to pair comparison, three-choice oddity problems, or matching from small samples. The first is the emphasis placed by differentiation theory on the relational nature of distinctive features. It is only in a particular context that the distinctions/

distinctions between members of the stimulus set in terms of their feature patterns become relevant; and if non-random confusion errors are anticipated it seems important to provide the opportunity for regularity in the properties of the stimulus set to be detected and used. This is linked, secondly, to the way in which "same" is understood; the limits of the subject's comprehension are most directly tested by observations of the sets of figures which he defines as being "the same." A relevant consideration here is the finding (Bryant, 1969), that children even somewhat older than this group have great difficulty in successive discrimination problems except under specially-arranged conditions of presentation. The inference from this is that regularities in a stimulus set would not be detected or used if the whole set was not available at one time. Finally there is the practical problem of number of stimulus presentations. An economical method of data-gathering is essential with small children, and this would not have been possible given the amount of material to be presented if anything other than a matching technique had been used.

A further consideration arose during pilot testing of the stimulus material, part of which was concerned with the feasibility of using pair comparison presentation. It was found that, particularly with figures differing only in orientation, there was a tendency for children between three and four years of age to label the figure and to base their judgements on the label. The question asked was "Is this picture just the same as this one?" accompanied

by/

by appropriate pointing by the experimenter. For example, when faced with two ellipses in different orientations a typical response to this question would be, "That's an egg, and that's an egg", or "Both eggs", rather than a direct "yes" and "no". Although similar labelling occurred at times with the larger arrays finally used in the experiment proper, particularly with the younger children, it was much easier to get the child to accept the idea that he was being asked something other than whether all the pictures were of the same thing.

The three sets of stimulus material shown in Fig. II were chosen. Series I, where orientation differences alone are used, represented a set of problems of increasing difficulty as measured by the number of errors made during pilot testing. It was intended to be the main index of change over time, as shape discrimination, as would be expected, was found to be relatively simpler even for younger subjects. Studying orientation separately implies disagreement with the point of view expressed by Wohlwill and Weiner (1964) to the effect that orientation discrimination is merely a special case of the more general issue of form discrimination. The lack of any immediately obvious distinguishing features in the former case, linked to the questions of the aspects used in differentiating among such material, provides some a priori grounds for accepting a qualitative difference between the two types of situation.

Series II and III were both intended to provide information on shape discrimination in relation to both shape and orientation differences, but from distinct points of view. Series II provided as much variation of attributes as possible,/

possible, while retaining an array of manageable size, to see what kinds of confusion errors were made when figures differed in shape, orientation, number of sides, and in being open or closed. Straight vs. curved discriminations were not tested, as children in our subjects' age group never confused rectilinear and curvilinear figures during pilot testing. With series III a more systematic approach to the same problem was attempted; apart from the two closed figures, retained as a further check on Series II findings, the choice set varied only in orientation (two values) and degree of openness (three values). The object here was to see if this minimum regularity in stimulus attributes would be reflected in any systematic fashion in response patterns.

Subjects

Twenty-three children were tested twice, at mean ages three years eight months (range 3-4 to 4-0) and four years nine months. There were twelve male and eleven female subjects; in addition one boy, who was not available at the second occasion, is included in the results of the first testing. The two testing sessions are referred to hereafter as pretest and post-test respectively.

Presentation

The figures comprising the three series were drawn in black ink on three-inch square white cards, which were then covered with washable plastic. During each session the experimenter and subject sat at a small table, appropriate/

appropriate to the size of the subject, with the experimenter on the subject's right. Series I problems were presented one at a time, the experimenter laying out the comparison array as shown in Fig. II with a duplicate of the standard figure just below the centre of the array. The experimenter pointed to the standard, told the subject to have a good look at it, and then said, "See if you can find a picture up here that's the same as this one," running his finger slowly along the line of comparison figures as he did so and pointing again at the standard at the words "... this one". Three different problems were used to introduce the task; with these, errors were explained and correct matches indicated. This was not done with the experimental problems; with these the experimenter made general approving comments, for example, "Good, that's the way", irregularly and no matter what the response was. The arrays for Series II and III were laid out as shown in Fig. II, duplicate copies of the standard being placed in the gap in the bottom row. Instructions were the same as for Series I, but with no introductory examples, and no correction of errors. Order of presentation of the standard figures was the same for all subjects, and bore no relation to the order of figures in the array. The three series were presented in the order of their numbering, problems 3 and 4 of Series I being presented again after Series III with the choice set laid out in the alternative order shown.

All subjects were quite happy to "come and play games" or "look at some puzzles". Time for a full session varied between fifteen and twenty minutes,/

minutes, depending on the amount of time taken up by general conversation between problems, but care was taken to see that subjects attended to each one as it was presented. No restrictions were placed on the amount of comment on the material or the number of matches made in any one problem; the experimenter did not respond to the former situation, and only queried the subject in the latter if it was said that all the figures were the same. In this instance he was asked if there was one which was just the same as the standard; again a multiple response was accepted if it was made.

RESULTS

Detailed figures are available in the paper by Taylor and Wales (1970), presented as an addendum to this dissertation. Only summary statistics are given here. Each series will be considered singly; this is followed by an overview of performance in all three series in terms of relevant common factors.

Series I

In the pretest two subjects gave the figure in position 1 in the array as the match to the standard in all problems; their results are omitted from the analysis.

Looking first at number of correct responses, there is considerable improvement over time in both absolute and proportional terms, as can be seen/

seen from Table I. The difference between correct choices on the two occasions is highly significant by the Wilcoxon matched-pairs signed-rank test (N=20; T=32.5; $p < .005$ (one-tailed)). The distribution of choices over position in the array for each problem also changes between pre- and post-test, as shown by the results of the Kolmogorov-Smirnov one-sample test; three of the fourteen distribution (of both first and total choices in seven problems, two being repeats) depart significantly from randomness in the pretest, whereas six do so in the post-test. The force of these findings is somewhat reduced by the presence of a strong position response bias. This is shown by comparing the number of times a figure in a given position in the array is wrongly chosen as a match to the standard with the number of times the choice of a figure in that position would have been correct (Table III, Taylor and Wales).

<u>Table I</u>

	Total correct first responses	Total correct responses	Total responses
Pretest	45 (20%)	65 (29%)	226
Post-test	80 (29%)	110 (39%)	280
	Series I summary results		

The order of preference for positions is 4, 3, 1, 6, 2, 5, with the first two being much more likely to be chosen than any of the others. Even taking this/

this into consideration, however, it is evident that the ability to make correct matches on the basis of orientation alone has increased considerably in these subjects. The increase in total number of responses requires comment; this is deferred until the results for the other two series have been presented and a more general consideration of the use of stimulus information is made.

Series II

Performance here was good in the pretest and improved in the post-test, as shown by the increased number of correct responses, the drop in total number of responses, and the relatively small number of confusions. (Table II). The change in the first of these is again highly significant on the Wilcoxon test: N=16: T=2.5: $p < .005$ (one-tailed).

Table II

	Total correct responses		Total responses
Pretest	136	(51%)	268
Post-test	171	(76%)	226
	Series II summary results		

A multidimensional scaling analysis of the pre- and post-test confusion matrices was carried out for the results of both Series II and III.

The/

The object of this technique is to find a spatial configuration for the stimulus set, in as small a number of dimensions as possible, in which the judged similarity of the members of the set to each other is reflected in their proximity in the configuration. As used here, the assumption underlying its application is that the more often two figures are confused the more similar they are seen to be. Essentially the procedure involved is an iterative one which finds the best-fitting monotonic relationship between degree of similarity (number of confusion errors) and distance. A measure of departure from monotonicity is given; in the version used here this value is called Delta. The smaller this is, the better the obtained configuration as a representation of the data. A discussion of the place of multidimensional scaling models and methods in relation to data of this nature is presented in Chapter 3; here it is only necessary to state that the method used here was developed by J. Doran (see Doran and Hodson (1966)), based on principles adumbrated by Shepard (1962a, b) and Kruskal (1964a, b).

The configurations for pre- and post-test results on Series II are given in Fig. III. Two dimensions give a perfectly satisfactory representation of the data. The axes of the plot merely provide a (normalised) frame of reference, and consequently are unlabelled. The features used in differentiating among the figures can be seen quite clearly; indeed in the post-test the distinctions are more definite than they at first sight appear to be, because of the small number of confusions made. The only changes between results on the two occasions can be seen to be of degree rather than kind. The three pairs/

pairs of figures differing only in orientation are well separated, while each member of each pair is close to the other (1 and 4, 2 and 7, 3 and 6). There are two main lines of division, between open and closed figures and three- and four-sided ones; the apparent violation of this in the position of figure 5 in the post-test configuration is due more to the lack of errors made in matches involving it, with twenty-three out of twenty-five being correct. The two "unique" figures, 5 and 8, while being most closely associated with the appropriate clusters (respectively open figures and closed four-sided ones) are sufficiently distant from them to indicate that features differentiating them are being detected and used to a fair extent. The highest proportion of correct matches was made to these two figures at both times. Figure 8 was the only one which in the post-test was matched to, and had matched to it, a non-identical figure differing in some feature other than orientation. Several subjects made comments along the lines of "That's a wee bit the same" or "That's the same but bigger" when comparing 2 and 8, thus indicating that they were aware of their difference in appearance but were uncertain as to whether or not they should call them "the same". In the pretest figure 8 was matched to several other figures a relatively large number of times; there was no recognisably systematic reason for this and the most plausible account of its popularity would seem to be a position preference effect comparable to that already described for Series I.

With the experimental method used here, where every member of the stimulus set is compared with the whole set, the full confusion matrix is generated/

generated rather than the half-matrix which (as will be described in the next chapter) is normally taken as the starting-point for the analysis of similarities data. To anticipate somewhat, it is a virtually universal assumption in discussions of similarity that it is a symmetric relation, that is, if A is similar to B, then B is similar to A. It follows that the corresponding entries in the two halves of a confusion matrix are expected to be identical, at least within the limits of experimental error. This is not the case here. There are quite substantial discrepancies between the number of times figure A is matched to figure B and the number of matches of B to A. The presence and size of such asymmetries is determined by the total number of matches made to each of the two figures involved, and the proportion of these that are correct. Clearly if 50% of the matches made to one figure are correct while 75% are to the other, the distributions of the erroneous responses would be expected to be discrepant. Thus a comparison of the number of correct responses made to each figure as standard should give as much information as studying the asymmetries themselves. It has to be assumed before taking such a step that such differences in response totals are not due to such extraneous factors as position preference.

Omitting figure 8 from consideration, in view of the probable position bias affecting responses involving it, six out of the seven asymmetries found (three on each occasion) occur between pairs of figures differing only in orientation. In the pretest, an average of 9% fewer correct matches are made to figures 1 and 3 than to figures 4 and 6, the rotated versions of figures 1 and 3; there/

there were 2% fewer matches to figure 7, the rotated square, than to figure 2. A similar pattern is apparent in the post-test; an average of 14% more correct matches are made to figures 4 and 7 than to 1 and 2, while only 7% fewer are made to 6 than to 3. Without wishing to make too much of this incidental finding, it does suggest that a form which can be coded as "x-plus", e.g. "square plus rotation", on the analogy of Woodworth's notion of schema-plus-correction, in terms of the context in which it appears, is more readily detected than a simpler version.

Finally, at this point the two "deviant" subjects as defined by the results of the classification experiment mentioned at the end of the previous chapter should be considered. Their pretest performance on this series, but not the others, mirrored that on the classification task. For one, "same in some way" was clearly equivalent to "identical in all respects"; he made no errors, of either confusion or omission, here. The other, in contrast, interpreted "same in some way" to mean non-identical; she got every item on this series wrong, apparently following a rule of her own to the effect that if you can't find a figure which is part of the standard, then find one of which the standard is a part. The permissible conclusion would seem to be that there is at least partial equivalence of performance across tasks, but at this stage it is not possible to say under what conditions such comparability can be expected.

Series III

Overall performance is less good here than in Series II, but nonetheless improvement over time is demonstrated in terms of the same criteria, with an/

an equivalent level of significance for the difference between numbers of correct responses being yielded by the Wilcoxon test. Table III summarises the results. The multidimensional scaling analysis shows that three dimensions are required to represent the data adequately. The configurations are given in Fig. IV. The two closed figures, 3 and 6, are well separated from the rest in both plots, again demonstrating the lack of confusability between open and closed figures. In the pretest, of the remaining six figures, 1, 2 and 4, with gaps uppermost, project in the opposite direction to 5, 7 and 8, with gaps at the side. Size of gap is reflected in the relative proximity of 2 and 7, least open, and 1 and 8, most open; 4 and 5 are more widely separated. This can be taken as indicating greater uncertainty in comparisons involving this last pair. Further evidence of this is given by the order of frequency in which the figures are matched to each other as standards. For pairs 1 and 8, and 2 and 7, when one member of the pair is the standard, the other is the figure which is most frequently incorrectly matched to it; this is not the case for 4 and 5, where figure 8 is more often given as being the same as both than is any other.

Table III

	Total correct responses	Total responses
Pretest	118 (45%)	259
Post-test	161 (64%)	249

Series III summary results

Thus, there is evidence for the use of both shape and orientation as criteria for differentiation at the younger age, in the making of judgements which, at least with the extreme values of the attributes available here, can to some extent be described as systematic. Nevertheless the relatively low number of correct responses together with the wide range of figures matched to any given standard shows that at best this is only rudimentary.

On the criteria of number of correct responses and distribution of erroneous matches over the available choice set, performance in the post-test shows greater specificity in the use of attributes. There is still considerable room for error, however, as can be seen from the scaling configuration. There are no clearly-marked clusters. The figure pairs, 1 and 8, 2 and 7, 4 and 5 each project in the same direction from the main plane of the plot, mirroring the fact that the figure most commonly confused with a given standard is the one which is the same "shape" as it but in a different orientation. At the same time figures differing in "shape" from the standard are also likely to be confused with it, but to a lesser extent. Table IV helps to clarify the situation. Figures differing from a standard in both shape and orientation are less likely to be matched to it in both the pre- and post-test, but the tendency is more marked on the latter occasion. Where there is only one attribute differentiating the figures, as many errors occur in the post-test as in the pre-test when this is orientation, but shape differences are more likely to be detected by the time of the second testing. It therefore seems that the relatively greater regularity of response is due firstly to the smaller number of incorrect matches and secondly to an increased refinement/

refinement in the use of available stimulus information.

There is a larger number of asymmetries here than in Series II, but no regularity even of the rather loose kind described there is apparent. It may be that the features marking variation in this stimulus set are less codable; but it is equally possible that these findings are fortuitous and are the outcome of experimentally irrelevant factors such as position bias or random error.

Table IV

	Dimensions of difference between Standard and comparison figures		
	Orientation	Shape	Both
Pretest	18.0	11.0	6.7
Post-test	18.3	7.5	1.8

Mean percentage errors on Series III as a function of number of attributes differentiating stimulus figures.

However, with experimental material such as that used here, where stimulus attributes are clearly not of equal value when it comes to differentiating among the members of the set, the presence of a feature which to some extent is not used in discriminating between them might well produce such an effect; whereas "X" may be considered to be to some extent similar to/

to "X-plus", the reverse is not the case. This suggestion is explored more fully in Experiment 3, where orientation differences are not present in one of the stimulus sets used, and where it may be assumed that a more equal weighting is given to the attributes comparing the set than would seem to be the case here.

The order of difficulty of the three Series, in terms of proportion of correct responses, is II, III, I, in both pre- and post-tests. This would be expected in view of the relatively greater difficulty of orientation discrimination for children of an age comparable to our subjects, and of the greater variety of attributes marking differences between the figures in Series II in comparison with Series III. What was not anticipated is that the order of difference in number of total responses between pre- and post-test is identical; there is a fall in Series II, a somewhat smaller one in Series III, and a not inconsiderable rise in Series I. The amount of increase here is not equivalent on all problems; it is most noticeable in problems 3 and 4 and their repeats, while in contrast number of responses to problems 1 and 2 drop fractionally, Inspection of the distribution for problems 1, 3 and 4 suggests that global factors of similarity between the stimulus figures, such as obliquity, are being detected; and that this is in turn related to the 'simplicity' of the problem, simplicity being defined in terms of proportion of correct responses. Thus on problem 1 in the pre-test, 15 of the 27 responses are correct; of the remainder, 6 are to the figure in position 3. This is the only other vertically-orientated figure/

figure in the array. By the time of the post-test performance on this problem has improved to the extent that this kind of analysis is no longer relevant; 22 out of 25 responses are correct. Taking problems 3 and 4 at the post-test, in the former the figure in position 2 (position 5 in the second presentation), the only "vertical" figure in an otherwise "horizontal" array, is chosen four times out of a total of 100 responses for the two, presentations, and in the latter, figures in positions 1 and 3 (3 and 2 respectively on 4R), the two non-oblique figures in the array, are picked eight times out of 94 responses. The standards in these cases fall in the more frequently chosen class of comparison figures. Since the directionality of the figures in problems 2 and 5 is virtually nil in comparison with the others, there is no point in examining the distribution of responses there from this point of view.

These findings, and their context, suggest an interaction over time between accuracy of response and number of responses made. Looking at the more difficult problems in Series I (3 and 4) at the pretest, it is clear that response is either random or based on position preference; this can be taken as representing the lowest level of performance. On problem 1 at the same time, on the other hand, some regularity is apparent, in that more than half the responses are correct, and that the largest number of errors is made with a figure which can be described as "more similar" to the standard than the others, in terms of their sharing a common orientation. (This figure is also in position 3, the second most commonly chosen position in the array; it/

it will be shown that the chances of position preference affecting this finding are small). Here a larger number of responses are made than on the post-test, where errors are minimal, as in the number of additional responses. At this time, the response pattern for problems 3 and 4 has changed, and corresponds more to that for problem 1 on the pre-test; there is a larger number of correct responses, and again some regularity is apparent.

Thus it seems that there are three distinct stages or levels of development of performance in a form discrimination situation apparent in these results. The first is that of virtual non-competence, where a small number of responses is given which are based at least as much on extraneous factors such as position preference as on stimulus attributes. At the second level number of responses increases, as gross features of similarity come to be detected; at the same time the number of correct responses also goes up, if only because of the increased chances of getting the right answer when more answers are given. And finally, the stimulus information defining the unique correct match is picked up, with a corresponding increase in correct responses and decrease in number of multiple responses (i.e. more than one comparison figure matched to the standard). Obviously level of performance at any one time will vary with differences in both problems and subjects; nevertheless, we would expect reasonably steady progress through these stages over time, with little if any regression to lower levels.

Subjects were classified, in terms of their performance on all Series, into/

into three categories. The second one described here was originally split into three sub-divisions (Taylor and Wales, 1970, Table IX): this has not been done here for reasons of simplicity of presentation. Criteria for inclusion in the lowest category (A) were no more than 25% multiple responses and no more than 25% correct responses, and for the highest (C) no more than 25% multiple responses and no less than 75% correct responses. The second category (B) accommodated all subjects between these extremes. It can be seen from Table V that there is a steady upward trend over time. This can be tested by calculating chi-square for each Series separately, giving a 3 x 2 contingency table for stages by occasions.

Table V

Classification of Subjects by level of performance on all three series in pre- and post-tests

Stage		Series I	Series II	Series III
A	Pretest	11	4	4
	Post-test	3	0	0
B	Pretest	12	12	14
	Post-test	16	10	11
C	Pretest	0	7	5
	Post-test	4	13	12

The/

The values are: Series I, 9.14; Series II, 5.98; Series III, 7.24. With two degrees of freedom a value of 5.99 is significant at the 0.05 level. Improvement in performance in terms of this system of classification is therefore quite firmly established. Table VI shows that there is relatively little falling back. The greater number of subjects who show no change on Series II and III reflects the fact that more of them had achieved the highest level of performance in the pre-test.

Table VI

Numbers of subjects failing to show improvement over time

	Series I	Series II	Series III
Down	3	2	3
No change	4	10	8

The variable difficulty of different types of problem at any time can also be tested, again using chi-square, with the data contained in Table V. Here two 3 x 3 contingency tables are formed, of stages by Series, one for each occasion. With four degrees of freedom, a value of 9.49 or greater is required to reach the 0.05 level; the obtained values for pre- and post-test results are respectively 11.90 and 12.75.

The evidence for the changes in strategy described above is fairly good.

Before/

Before it can be fully accepted, however, it must be shown that subjects in category B, those who are taken to be using global criteria of similarity in making their many responses, are in fact using stimulus features as the basis for their choices rather than position preference. This can be done by looking at the number of times subjects in category B described one or other of the comparison figures in problems 3 and 4, Series 1, which have been described above as being less similar to the standard in gross terms, as a match to it. In problems 4 and 4R one of these figures was in position 3, the second most commonly chosen one in both the pre- and post-test. Table VII shows the number of responses to these "less similar" figures in relation to the total number of responses made by subjects in category B.

Table VII

Influence of gross features of similarity on responses of subjects in Category B

Problem	No. subjects making multiple responses	Total responses	Position of "less similar" figures	No. responses to "less similar figures
3	17	43	2	1
3R	13	30	5	2
4	14	34	1,3	1
4R	10	28	3,2	2

It is quite clear that if position preference is affecting their judgements, it is certainly not predominant, and that global features of similarity are much more/

more likely to be responsible for the distribution shown.

DISCUSSION

It is clear that any description of the uses of stimulus attributes in this experiment has to take into consideration the level at which the subject is performing. The evidence for the developmental sequence of strategies which has been inferred from these results is sufficiently compelling for it to be given due weight in any discussion of form discrimination; it is perhaps the most important finding to come out of this study, inasmuch as it gives some indication of the kind of changes involved as the child progresses to the level of full competence in handling perceptual tasks of this nature.

The transition from the first to the second level is analogous to a phenomenon noted in a different context by Donaldson (1963), an increase in error following the partial development of new abilities in the child. Clearly the analogy is not perfect, in that errors decrease here, but a similar process would seem to be involved. It appears that the child comes to realise (not in any conscious sense) that more factors have to be taken into account in a problem solving situation than had hitherto been considered, but that the capacity to process these additional aspects takes rather longer to develop. As a result there is a qualitative change in performance, which may be reflected in a greater number of erroneous responses, or, as is the case here, a greater number of responses and hence of correct ones. In both cases the effect comes from a greater awareness of the features of the situation/

situation which are relevant to the successful completion of the task imposed, but an awareness which is not as yet sufficiently educated to be able to filter out irrelevant aspects. This level of competence is represented in the third stage described above, where most responses are correct and few confusions are made.

Further clarification of the issues involved requires a brief consideration of the different ways in which "same" can be used. That given here is modelled on the presentation by Donaldson and Wales (1968). The first sense is that where "same" refers to the identity of one object over time, as in "I still have the same car." Three different senses of "same" can be distinguished when two or more objects are compared at any one time; where the objects are identical in respect of all observable attributes, where they are identical with respect to at least one observable attribute, but also differ with respect to at least one, and finally where they are classified as "the same" in terms of non-observable attributes, e.g. "Women are all the same."

Only the two cases relating to observable attributes are relevant here. It is clear, both from the results of the classification task mentioned above, and from the performance of subjects classified at stage B in this experiment, that pre-school children are quite well able to understand the relation of partial identity and that they can apply it systematically in some situations. Relatively few, on the other hand, are able to work in terms of complete identity; even those who have reached stage C make some confusion errors. However, the non-equivalence of performance with different types of stimulus material on both testing/

testing occasions indicates that the capacity to use this understanding is limited by the complexity, in some sense, of the task involved. The child appears to have an adequate conception of similarity, broader perhaps than that of an adult but not grossly deviant from it. (cf. the difficulty some subjects had, having described the square and rectangle of Series II as "a wee bit the same", in deciding whether or not to classify them as "same" or "not-same"), but is unable to apply it in certain situations because he is unable to find the necessary information in the stimuli. Thus, the pattern typical of stage A appears, with essentially irrelevant factors guiding responses. The increase in number of responses at stage B, where the choices made can be shown to reflect some degree of similarity to the standard figure, demonstrates both the development of the realisation that there is a principled basis for answering the question "Is this the same as this?" and a lack of specificity in utilising the stimulus information which provides this basis.

As far as the question of the interaction between discrimination performance and the use of the relation of similarity is concerned, it is clear that changes in response patterns with time reflect increasing sophistication in the detection of the stimulus information in terms of which comparative judgements are made rather than any alteration in the child's understanding or application of the relation. This statement only applies, obviously, once the concept has become established in the child, which can be taken to be at a younger age than that of these subjects given the results of the classification task. The following conclusions can be drawn about the other two issues mentioned/

mentioned as a focus of this experiment, the use made of stimulus features which are criterial for differentiating between the figures and whether or not errors show any systematic pattern. The results from Series II and III show that even at three-and-a-half children can discriminate between figures of different shape with considerable success, and that there are quite distinct class boundaries to the confusion errors that they do make. The number of sides to a figure, and whether or not it is open or closed, are used more often than orientation as differentiating criteria. With a more systematic set of stimulus figures, as in Series III, it can be demonstrated that fewer confusion errors will be made the greater the number of attributes there are marking the difference between figures.

In both these Series, the emphasis on the use of shape as a determinant of similarity rather than orientation, indicates that the former has greater value as a distinguishing feature. From Series I results, we know that orientation differences can be detected, but it appears that where other attributes vary as well, this is much less likely to occur. The work of Vurpillot (1968) is relevant here. She extended the Russian findings (Zaporozhets, 1964) on the irregularity of visual scanning in children up to the age of five by demonstrating that not only are systematic scanning strategies, where each part or attribute of a figure is compared with the corresponding feature of another, slow to develop, but that younger children also fail to search the stimuli fully and so make their comparative judgements on the basis of only partial information. A similar state of affairs seems to hold for the results reported/

reported here; once one common feature is detected in a pair of figures, the search for further confirming or disconfirming information is discontinued.

Where orientation is the only distinguishing feature present in a stimulus set, it is more difficult for the child to detect the criterial differences; the poorer level of performance on Series I shows this quite clearly. There is fair evidence that gross features of similarity such as obliquity are detected as the child takes its first steps along the road to systematic error-free responding; however, it is only with stimulus figures where directionality is clearly marked that this can be expected to be demonstrated. This finding is more in accord with Rosenblith's (1965) comment that figures may "behave" differently under similar transformations, that different error patterns may be obtained depending on the relations holding between the standard and the choice set, than with the suggestion made by Gibson et al (1962), that transformation type alone is a more useful predictor of error. It has been seen that a 90° rotation of a horizontally or vertically oriented standard figure (problems 1, 3 and 3R, Series I) is less likely to be confused with it than is a comparison figure rotated 90° from an oblique standard (problems 4 and 4R, Series I).

To sum up the findings of this experiment, therefore, it can be said that at least with children of the age of these subjects, errors in a matching-from-sample discrimination task are due more to failure to detect criterial distinctive features and to make an exhaustive search for these than to their using "same" in an incorrect fashion; the reference of the term to a relation between two or more figures is understood, and it is interpreted as indicating at least partial identity between figures. Improvement of performance over time/

time is the result of increased specificity in the detection and use of distinctive stimulus features. Individual differences in attainment of different levels are wide, but the sequence is quite steady; following the initial essentially random stages of responding, where features of the total experimental situation are as likely to determine responses as stimulus attributes, a stage is reached where wide but systematic classes of similar figures are defined, the sense of similarity here being that of partial identity. This gives way to the final stage where correct and unique matches are made, with criterial distinctive features being detected and irrelevant ones ignored.

Comparing these results with those of the studies reviewed earlier leads to the suggestion that conclusions about the factors most likely to influence form discrimination are going to depend very much on the experimental context in which the question is studied. Whatever the relation between the standard and comparison figure(s), once the child has passed the first stage described here his response is going to be systematic to some extent. However, it is only in the situation where a large comparison array is available and an unrestricted range of responses is open to him that this will be clearly seen, unless he has reached the final stage of making correct unique matches. Transformation type is less important than the nature of the standard figure in predicting errors, but again this is only fully apparent in a relatively unstructured situation. As to the type of distinctive feature that is most likely to affect discrimination one way or the other, there seems little point in being dogmatic. Orientation is more difficult to use than one which alters the shape of a figure, but/

but nevertheless can be and is detected. In general, it seems fair to say that any difference between figures can provide a basis for differentiation, if it can be detected, and that its detection is largely a function of the child's ability to make a systematic search of the stimulus figures.

The question of the use of stimulus features will be examined much more fully in the next series of experiments to be described, along with other issues. Before this however, it is necessary to discuss certain theoretical and methodological problems raised by this study.

Chapter 3

The Concept of Similarity

The issues to be discussed here can be summarised briefly in the form of three questions: What is being discriminated? How is it being discriminated? How adequately are the discriminations being made? These somewhat clumsy formulations can be more accurately rephrased in terms of the discussion of similarity as an abstract concept given by Donaldson and Wales (1968), where two or more entities are described as similar if they are partially or totally equivalent with respect to their observable attributes or features. The first question then becomes, What are the elements of the stimulus set presented to the subject, which are defined or definable by the experimenter? The problem posed here is that of the specification of stimulus properties in advance of, or at least independently of, the subject's evaluation of them. The second question is intended to emphasise the judgemental aspect of the subject's task, and in fact breaks down into two distinct parts. The first of these includes the notion, previously discussed at some length, of the subject's understanding and use of "same"; it refers to the decision rule on the basis of which he assigns members of the choice or comparison set of stimulus figures to the categories "Match standard"/ "do not match standard." The second part relates to the detection of salient stimulus features, on the basis of which the decision is made. Again, two separable issues are involved in the third question. There is the methodological one of the type of analysis appropriate to/

to the examination of similarity judgements, and the assumption involved in the various alternatives; the other is concerned with the adequacy of the subject's representation of the relations holding between the members of the stimulus set. The two parts of the second question, and the latter part of the third, are quite obviously to be answered empirically, and indeed they represent the basic concerns of this thesis. Consideration of the other points, however, is a necessary preliminary to their experimental examination.

The relevance of these questions to the study of form discrimination can be brought out more clearly if we consider the situation in more general terms. Some years ago Galanter (1956) presented an analysis of the problem of "... (studying) the order of the organism's experience as determined by a study of behaviour." He started by pointing out that the content of experience cannot, in an objective psychology, be communicated directly, only in terms of the structure or relationship between perceived (experienced) objects, as in such statements as "red looks more like orange than green does." Taking as an example the subject matter of classical psychophysics, Galanter points out that three things must be considered. One is the distribution of physical energy underlying a particular area of experience, such as sound pressure, which can be ordered in terms of physical measurements; another is the experience resulting from exposure to such physically-defined events, which can be seen under appropriate conditions of examination to reflect to some extent the order inherent in them. Thirdly, there is the function which maps the one order into the other, although not in any one-to-one sense.

The/

The central problem for Galanter is the construction of a map for each sense realm which will assign a unique position to each element in the realm in such a way that increased likeness of experience of two elements will be reflected in their closer proximity in the map. It is clear that what is being attempted here is a very general, context-free, construction; many experiments carried out since Galanter's paper was published indicate that this is not justifiable, for example, Garner (1966). Galanter's formulation, however, is applicable to the problems with which we are faced here. Experienced order, or, to revert to more appropriate terminology, the perceived structure of a stimulus set, has to be related to its physical structure in such a way that we are able to make some inferences about the way in which the latter has been mapped into the former and so come to some conclusions about the nature of the subject's information processing capabilities.

What we require, in fact, is a normative model in the sense described by Garner and Morton (1969). The purpose of such a model, which is usually expressed in mathematical form, is the establishment of a baseline against which performance can be compared. This emphasis is important; the correctness or otherwise of the model is not to be evaluated only in terms of the adequacy with which it accounts for the data, but also in the extent to which it reflects the theoretical points which the user intends it to reflect. Thus we must try to find a model which will provide an explicit comparison between the perceived and physical structure of the stimulus set

in/

in such a way that discrepancies can be accounted for in terms of the processes used by the subjects in deriving the latter from the former. From a more empirical point of view the importance of providing a principled description of the stimulus set is evident in the quotation taken from E. Gibson (1969), given on p.1 of the Introduction. Since we are concerned with "... the growth of ability to detect regularity, order and structure," it is necessary to provide suitable opportunities for such ability to be expressed. Experiment 1 demonstrated that, with those subjects who were capable of responding at anything beyond a random level, there were quite clear-cut classes of figures which were confused with each other; open figures were rarely matched with closed ones, for example, and where the stimulus set was defined by two attributes only, figures differing from the standard on both attributes were less often confused with it than were those differing on one. Some regularity of response pattern, mirroring the features of the stimuli, was therefore apparent. However, the lack of any systematic relation between the stimulus classes in the material of Series II, and the use of an attribute (orientation) which is associated with particular difficulties in discrimination in Series III precludes a complete assessment of the ability of young children to detect and process class relations within these stimulus sets when presented as a discrimination task.

The notion of structure as used here can again best be presented in the terms used by Imai and Garner (1968); "... the relations between stimuli or the attributes forming a set of stimuli, rather than the isolated properties of the/

the stimuli themselves." E.J. Gibson's discussion of distinctive features emphasises their relational character; describing a feature as distinctive makes sense only in the context firstly of other features which do not have this role and secondly of a set of stimuli which can be partitioned on the basis of the particular feature. In fact it is a moot point as to whether a feature, or a set of features, is accurately described as relational; it is the description of the elements of the stimulus set, in terms of a feature list, which is relational and is important here. We can talk of a number of different types of structural relations. There is the structure of the total stimulus set, and that of the sub-sets into which it can be divided, as defined by the experimenter; and there is the structure of, in this context, the sub-set of the total stimulus set which the subject defines as equivalent in terms of his confusion errors. It is the comparison between the structure of the subject's sub-set and that of the sub-sets available to him in the total set that provides the basis for inferences about his perceptual functioning.

As Imai and Garner go on to point out, structure can be defined in either metric or non-metric terms. Strictly speaking the latter should be apparent only in qualitative differences between stimuli; that is stimuli would be either identical or non-identical, and if the latter, then it should not be possible to make any further statements on the nature or degree of the difference between them. In practice, however, this is not feasible, as even the counting of numbers of attributes in common can be considered as a simple/

simple kind of quantification. Imai and Garner's solution to this problem was to demonstrate empirically that, within a set of nine stimuli generated by two attributes (shape and colour) with three levels of each (circle, square, triangle: red, blue, yellow),perceived similarity depended only on the number of attributes by which stimuli differed. Thus two stimuli differing only in colour were judged to be as similar as two stimuli differing only in shape, and both pairs were judged to be more similar than was a pair differing in both shape and colour. The distinction of importance thus comes down to that between continuous attributes, where differences in level or value of the attribute can be expressed in terms of a numerical scale, and discrete attributes, where differences in level are of kind rather than degree. Shape is an instance of the latter; there is no continuum connecting triangles, quadrilaterals and circles. The former presents greater difficulties, however. It is possible to think of a metric underlying the attribute of colour, particularly if we consider the three components of luminance (brightness), purity (saturation) and wavelength (hue), each of which has a distinct scale associated with both its radiant energy or physical aspect and its observable or psychological aspect. If, however, we extract three colours from widely separated areas of the colour solid these can be taken as discrete qualities, with the immediately-observab le differences between them making insignificant the fact that they stand on a common set of dimensions. Orientation is another such ambiguous attribute, inasmuch as it can be conceived of in either dimensional or class terms, in a way exactly/

exactly analogous to the case of colour.

A final point to be taken account of in choice of stimulus specification is the nature of the relationship between the attributes comprising the stimuli. A number of authors have commented on the distinction between unanalysable and analysable stimuli. Shepard (1964) describes "... those that are reacted to as homogeneous, unitary wholes, and those that tend to be analysed into perceptually distinct components or properties", and Torgerson (1965) talks of multidimensional attributes as contrasted with sets of stimuli varying on several distinct attributes. Colour is the standard example of the former class, while the latter refer to figures such as those used by Shepard, which varied in size of circle and in inclination of radius. This distinction is of more than theoretical interest; there is considerable empirical evidence, which will be reviewed at a later stage, that stimulus sets differing in this respect are processed in different ways.

With one exception (Brown and Goldstein, 1967) all the studies reviewed in Chapter 1 used stimulus material which varied in discrete rather than continuous attributes. Those where orientation discrimination alone was studied used at most a two-choice matching task; the conclusion drawn from the results was that particular types of rotation or reversal were more difficult to discriminate than others, and that the context provided by the stimulus display could affect the ease with which the figures could be differentiated. However, with a larger set of comparison figures, as was used in Series I of Experiment I here, the findings suggested that it

was/

was not the absolute nature of the difference between the standard and comparison figures which was important, but rather the relation between the features of both. A figure rotated 90° from the standard is more likely to be confused with it when both are oblique than when thettwo are aligned along the horizontal-vertical dimensions. Thus by increasing the amount of information (in the non-technical sense) available to the subject in the structure of the total set of stimuli to which he is responding, grounds are provided for extending the permissible inferences about the way in which they are processed.

Of the form discrimination studies cited, that by Birch and Lefford (1967) made use of a stimulus set with relatively little definable structure to it, much as in Series II of Experiment 1. The minimal number of errors made by their subjects, the most common one being a failure to detect, or at least to respond to, an orientation difference, made any analysis of response patterns in terms of confusion errors irrelevant. The two Gibson studies, however, produced more fruitful material. In both a higher degree of structure was available to the subjects, but of a different kind. Gibson et al (1962) used the same transformations of a number of different forms, and their conclusions were based on generalisations across results from stimulus sets which were connected by (ostensible) equivalence of difference between the standard and the comparison figures. Only within one transformation type, straight line to curve or vice versa, were the different levels, represented here by the number of lines which were varied in the comparison figure, related in any way; and results showed that the greater the/

the number of points of difference between the figures, the smaller was the likelihood of their being confused. Different degrees of rotation and types of reversal were shown to produce different error rates, but the inferences drawn have been shown to be accurate only to a limited degree, because of the restricted nature of the standard figures chosen (see Rosenblith's (1965) comments, p. 43 below).

However, the structure of such stimulus sets is limited by virtue of the fact that each comparison figure differs from the standard in terms of only one feature; with straight/curve and rotation or reversal transformations a number of values were provided, thus giving some indication of the connection between degree of difference between figures and their perceived similarity, but in no sense were there any links between the four major dimensions of difference in the comparison set. Thus relations between features, and their respective utility in a discrimination task, can be assessed only in terms of number of confusions of certain types of transformation with the standard figure at different ages.

The second study (Gibson et al, 1964) examined the discrimination of upper-case letters of the alphabet. The use of a list of bipolar features to describe the elements of the stimulus set produced a definition of its structural relations where the degree of similarity between any two elements was reflected in the number of features they had in common; here, therefore, structural relations within the total set are defined. The experimental findings indicated that the more similar letter pairs were, in terms of their common features,/

features, the more likely they were to be confused; a multidimensional scaling analysis of the obtained confusion matrices suggested that two features, straight-curve and relative obliqueness, were particularly important for discrimination. The intention of this study was to clarify the dimensions of difference between letters which are criterial for identifying them, and its outcome, showing good agreement between the structural description of the stimuli and the subjects' perception of it, indicates that the former was well-founded.

The Brown and Goldstein study is representative of an approabh to the problem of stimulus specification which has come to be known as form psychophysics. Much work has been done in this area and it will be considered in some detail. Its initial impetus came from an article by Attneave and Arnoult (1956), describing several methods for generating and quantifying random polygons. Stressing the importance of formal or relational factors in perception, they were concerned with systematising these in order that results from one experimental situation might be generalised not only to different subjects but also to new stimuli. They accepted, without argument, that the psychophysical paradigm is appropriate, although the difficulties raised by the multidimensional nature of shape (as contrasted with the usually unidimensional nature of stimuli used in psychophysical experiments), the variation in the number of dimensions needed to describe a shape as its complexity increases, and the lack of knowledge about the psychological relevance of such physically specified dimensions, were acknowledged. The problem/

problem of generalising from a given set of stimuli was broken down into two parts, one that of defining the stimulus domain, that is the set of stimulus figures to which results may be generalised, which is to be done in terms of the statistical parameters characterising it, and the other that of producing a sample which has ecological validity in Brunswik's sense, necessary if results are to be generalised to natural forms. Only the first of these was discussed, the second being taken as a special case of the more general issue.

The body of their paper presents a number of different rule systems for generating, on a random basis, polygons of different types together with a discussion of various ways of quantifying them. A stimulus domain is thus defined in terms of the set of rules used to produce the figures. The authors conclude, however, on a somewhat pessimistic note; "... it appears unlikely that any single system of physical measurement can be optimal for all psychophysical interactions ... there is no quick and easy way to determine which physical measurements have greatest psychological relevance: only experimentation can answer this question." (Attneave and Arnoult, 1956, p. 470.)

Much research was carried out following the appearance of this paper, with emphasis on the development of a viable psychophysics of form perception. However, a review of the literature, (Michels and Zusne, 1965) made it quite clear that the basic problems, to do with the most appropriate metric for quantification of visual form and the use of this in prediction, had been largely undisturbed. Since then an extensive programme, on this same topic, has been launched by D.R. Brown and his colleagues. The first paper in their series stated the thesis that "... the study of measurement and sampling of stimuli is propaedentic to further/

further development of a psychophysics of form perception" (Brown and Owen, 1967, p. 243), without any discussion of whether or not the psychophysical model is in fact the most appropriate one when applied to form perception. They selected as their stimulus population a large sample of random polygons, generated by Attneave and Arnoult's Method 1; 200 at each of five "sidedness levels" (4, 8, 12, 16 and 20 sides) were used. Two considerations dictated this choice of material, the more obvious being the popularity in previous studies of such stimulus sets. The other was that material of this nature varied both in type of attribute differentiating among figures and also in amount of each attribute present in each one. Eighty different measures were applied to the population; these were based on angle size, length of side, area, perimeter units, radial length, and co-ordinates. Moments of the distribution were calculated for these values. Values of the different measures across stimuli were correlated in an R-technique factor analysis, so that factors were defined in terms of clusters of measures rather than of stimulus figures.

It was found that the number of factors required to account for a given percentage of variance of the 80 measure variables increased with the number of sides in each stimulus sample, indicating that the number of independent measures required to achieve a given level of description is directly related to the number of sides of the figures. The nature of the factors was discussed using results from the quadrilateral and 12-sided figures, as typical of the remainder of the findings. However, the presentation was restricted to the first five factors, accounting respectively for 74% and 58% of the matrix variance for the two samples, on the ground/

ground that these accounted for a relatively large proportion of the variance, were meaningfully interpretable, and that the remaining factors in each case contributed small and approximately equivalent increments. These factors were labelled: compactness, being a measure of the dispersion of the perimeter or area of a figure from its centre of gravity; jaggedness, related to the proportion of acute interior angles; skewness of area and perimeter relative to the x and y axis. A large number of individual measures were correlated with each of these.

The authors also tackled the question of sampling variability. This was done by considering repeated samples of stimuli from a given population as providing estimates of the values of the statistical parameters defining the population; the stability of such estimates across samples thus provides an index of the variability, or lack of it, to be expected in drawing different samples. The method used was simply to compare the factor solutions at each sidedness level for two random-sub-samples of 100 shapes, thus indicating the extent to which relations among measures corresponded in the two sub-samples. Quadrilateral shapes showed a high degree of stability, but this fell off as the number of sides increased. It was concluded that the degree to which shapes were free to be unique increases as number of sides increases; as a result the correlation between measures decreases, and larger samples are required with figures of a greater number of sides to compensate for decreased stability.

Brown and Owen's study has been quoted at such length because of the thoroughness and rigour with which it was executed, and as the best available example of the psychophysical approach to the problems of form perception. In spite of this, their concluding statements, rather like those of Attneave and Arnoult,/

Arnoult, make depressing reading: "A reasonable attitude at this point in time seems to be that we do not have the quantity or type of data needed to arrive at final conclusions concerning a measurement system for shape." (Brown and Owen, 1967, p. 257). And they end their paper with a quotation from Arnoult, published in 1960, which they describe as still appropriate. In part, this reads, "The development of a psychophysics of form still lies far in the future .... only the barest beginning has been made..." The procedure they propose to adopt in order to refine the measurement of form is a circular one. They point out, almost as an afterthought, that aspects of similarity are involved, and that two approaches are possible in studying perception. Either the conformity of subjects' similarity estimates to a predefined stimulus domain can be assessed, or the domain can be specified post hoc in terms of an analysis of subjects' responses. Their iterative process combines both lines of attack, working back from experimental results to more psychologically relevant form measures.

One of the most noteworthy features of Brown and Owen's paper is the increasing generality apparent as the argument develops. Starting from highly specific physical measures, for example, the second areal moment of x, they end up with factors labelled in terms of their appearance. The second areal moment of x is one of the 29 measures with highest correlations on factor I, compactness, for four-sided figures; but there is no suggestion that any one of these is in any sense the best representative of the factor. The measures are, obviously, highly intercorrelated, but not perfectly so; thus each one contributes some unique information about the figures to which it is applied. Specificity has/

has been lost in two senses, one that the stimuli are no longer defined by any one measurement operation, the other that they are, of necessity, described in psychological terms, applied by the authors, which are applicable to the lowest common denominator of the outcomes of a number of physical operations. The one-to-one, or even several-to-one, correspondence between the physical and the psychological, between type and nature of stimulus and the response to it, which is the basis of sensory psychophysics, is replaced by a potentially many-one, or more likely many-several, relation.

Considering the proposed experimental programme for refining measurement operations in the light of the goal of developing a characterisation of a stimulus sample such that results can be generalised to new samples, another difficulty arises. It has been seen that, even within a given level of sidedness, the stability of the sample decreases as number of sides increases. The implication is that conclusions based on experiments using less than the whole set of 200 stimuli will vary in accuracy, depending on the number of sides to the figures, perhaps to the extent that different conclusions about perceptual functioning may be drawn from the use of different subsamples of a defined population. Again the potential for departure from the tightness of standard psychophysical relations is apparent.

The gist of these comments on the psychophysical approach to form perception is that the effort is misconceived. Two further points can be made here. One relates to the almost pathological attraction for psychology of the splendours of mathematical rigour. Schwartz (1962), writing on the role of mathematics in the formulation of physical theories, demonstrates quite clearly/

clearly how dangerous an ill-thought-out application of such methods can be in both the physical and social sciences. His basic argument is quite simply that mathematics, like computers, is single-minded, literal-minded and simple-minded, a most inappropriate set of attributes in a tool used by the scientist in attempting to deal with the ill-understood approximations to reality which are his stock-in-trade. A few quotations will make Schwartz' position clear; the implications for the present discussion should be obvious: "...it (is) essential, if mathematics is to be appropriately used in science, that the assumptions upon which mathematics is to elaborate be correctly chosen from a larger point of view, invisible to mathematics itself .... Mathematics is able to deal successfully only with the simplest of situations, more precisely, with a complex situation only to the extent that rare good fortune makes this complex situation hinge upon a few dominant single factors": "Related to this deficiency ....is the simple-mindedness of mathematics - its willingness .... to elaborate upon any idea, however absurd: to dress scientific brilliancies and scientific absurdities alike in the impressive uniform of formulae and theorems. Unfortunately, however, an absurdity in uniform is far more persuasive than an absurdity unclad."

The second point concerns the kind of misconception that is evident; an analogy with auditory experiments is the best way of doing this. In straight-forward psychophysical detection and recognition experiments in vision and audition the stimuli used have quite definite physical representations, in terms of wavelength, pitch, luminance, intensity, and so on. With auditory stimuli, however, as soon as some kind of higher-order structure or pattern

is/

is imposed on them the representation changes. The structure can be relatively trivial, for example, a short sequence of tones taking one of two possible values, or more complicated, for example, speech sounds; but experimenters using such materials do not start from a precise specification of the sound patterns involved (cf. Royer and Garner, 1966). The problem is too complex, perhaps because of the sequential nature of the organisation inherent in the stimuli, but such an approach is also irrelevant, in that more appropriate characterisations of the stimulus material are available. These can vary from the complexities of linguistic theory to the simplicity of the experimenter's intuition, the level to which Brown and Owen retreat in the description of their form measure factors. That organisation and structure of a type qualitatively distinct from that of radiant energy are involved in two-dimensional forms is accepted by all who work with them; it is therefore all the more surprising that a relatively uncritical acceptance of a methodology appropriate to the study of lower-order phenomena should be manifested. Brown and Owen state, explicitly, that aspects of similarity are involved in the type of problem they attempt to resolve; but they make no attempt to shift their frame of reference to take account of this, giving rather the impression that form psychophysics is an end in itself rather than an experimental tool.

At this point it seems fair to say that the system used for the formal specification of stimuli should provide a complete and exclusive description of/

of each member of the set in relation to every other one, so as to enable subjects to show as unambiguously as possible what features they are using and in what way they are using them. However, before examining further evidence relevant to the selection of criteria to be used in deciding which is the most appropriate form of representation of stimulus relations, some attention must be given to the methods available for the analysis of similarity data. In the area of experiment with which we are concerned, discrimination tasks using a matching response, the data can take two basic forms, latency of response or number of errors. The following comments are restricted to the latter only, partly because this was the measure used in the experiments reported here, but also and more importantly because the two are not equivalent. This stems from the fact that time is a continuous variable whereas match/mismatch judgements are not. In the hypothetical case of a series of stimuli of increasing similarity, it would be expected that this would be reflected even for individuals, in a corresponding increase in response latency; with matching errors, however, there would be, in the individual case, a cut-off point separating the perceived non-identical and identical stimuli. In this case a number of subjects would be required to produce a similarity order rather than a dichotomy. Furthermore, it is not immediately obvious what the equivalent in terms of response latency would be to a failure to match two identical stimuli. Standard statistical techniques can, of course, be applied, but these are not designed to bring out the kind of structural relations which are/

are being emphasised here, particularly where each member of a stimulus set has been compared with every other. Multidimensional scaling (MDS) is a recently-developed technique which is designed for just such problems; the assumptions and methods it involves, and the nature of the solutions provided by it, need to be considered in some detail.

The basic concept behind MDS is that similarity relations within a set of objects can be represented as distances in an M-dimensional space. The problem is to find a monotonic transformation of the N (N-1)/2 similarity measures, obtained from a set of size N, which will convert them into distances and so permit the construction of a configuration of the N points such that the more similar two points are judged to be, the closer they are in the configuration. A number of suggestions and techniques have been put forward, of which three will be discussed here.

The first line of approach (Torgerson, 1958) used a two-stage procedure, whereby an adaptation of standard one-dimensional scaling techniques was first applied to the similarity measures to convert them to distances; the second step was to derive the configuration whose interpoint distances corresponded to those resulting from the first operation. However, the use of a variance-dependent method, with the distributional assumptions entailed in such an approach, together with the stringent requirements that the relation between similarity measures and distance be linear and that the distances be Euclidean, imposes severe restrictions on the applicability of this model.

A different approach was initiated by Shepard (1962a, b); his procedure,

and/

and the development of it by Kruskal (1964a, b), have been by far the most widely used over the last decade. A version of it was used in the analysis of the results of Experiment I. Here no assumptions are made as to the form of the function relating similarity and distance. The only requirement is that it be monotonic, that is, that distance increases as similarity decreases; in practice, Shepard's technique attempts to find a configuration such that the rank order of the interpoint distance is the inverse of the similarity measures. Thus the vexed question of carrying out arithmetical operations on ordinal data is sidestepped. Different problems are now presented, however. One relates to the fact that it is always possible to embed N points in a space of N-1 dimensions. This would provide trivial solutions, in that no real reduction of the original data to a more comprehensible form is achieved. The number of dimensions defining the space can only be lessened at the expense of a departure from a strict monotonic relation between similarity measures and distance. The final solution is therefore a compromise between maximising the goodness of fit of the order of the interpoint distances to the order of similarity measures and minimising the number of dimensions required for the representation of the stimulus configuration. The procedure is iterative, in that an initial randomly-generated configuration in N-1 dimensions is adjusted until monotonicity is achieved. The number of dimensions is then decreased by one and the process repeated, if desired down to the one-dimensional case.

There are a number of points of difficulty associated with this approach.

The/

The first relates to the nature of the data. As has already been pointed out (Chapter 2) it is assumed that similarity is a symmetric and reflexive relation; the former is particularly necessary where a spatial model is being utilised, since the straight line distance between two points is a constant and therefore not subject to variation depending on the direction taken between them. If asymmetries occur in the data, as was the case in Experiment I, they must be removed, by averaging, before MDS techniques can be applied. Potentially important information may thus be lost unless such anomalies are considered in their own right.

The lack of any rigorous criterion for the evaluation of departure from strict monotonicity in the function relating similarity and distance is unfortunate. Certain values for the error term involved, be it discrepancies in the rank order in the Shepard approach or the residual variance of the monotone regression of distance on dissimilarity, called stress, in Kruskal's, have been suggested on an empirical basis as suitable cut-off prints, but there is no formal justification for this. Kruskal (1964a) has proposed that the appropriate number of dimensions is that beyond which stress falls comparatively little; this point would be indicated by an "elbow" in the plot of stress against number of dimensions. Thus the choice of final configuration is to a certain extent arbitrary. In practice there is little point in accepting a configuration in more than three dimensions, as this would make its representation virtually impossible, unless the departure from goodness of fit was so great as to be quite unacceptable.

Linked/

Linked to this is the question of interpretation. It is perhaps unfortunate that the name of the technique should incorporate the word "dimension". This appears to have suggested to certain authors (for example, Behrman and Brown, 1968) that the projections along individual axes of the plot produced by an MDS program are the most relevant part of the analysis. This is not the case, as can be seen by recalling what is involved in producing the configuration. This is derived from the inter-point distances which give the best approximation to the obtained similarity measures. It is only in the context of the whole configuration that the spatial relations between the stimulus points are fully manifest, since only then can the distance of each from each be accurately assessed. This is not to say that consideration of the individual dimensions is meaningless or wrong, simply that it should not be taken for granted that they are the most important aspects of the results. Behrman and Brown plotted each dimension of an MDS - derived configuration of 4-sided random polygons (taken from the sample described by Brown and Owen, 1967), all possible pairs of stimuli having been rated for similarity, against various physical measures of the forms. They found a good level of correlation between the first two dimensions and two of their measures, but not with the third dimension. The point here is that the attitude exemplified by this study fails to consider the possibility that subjects may not be responding to the structure the experimenter expects them to, and in addition overlooks the distinction between continuous and discrete variation, or, as Torgerson (1965) puts it, quantitative and qualitative variables. He presents a set of results demonstrating/

demonstrating conclusively that subjects judging stimulus sets which can be quite adequately described in terms of two quantitative dimensions, in this case width and degree of asymmetry, introduced a third aspect of variation in that they also considered the direction of the asymmetry. This aspect of their performance would have been lost if a two-dimensional configuration had been obtained and the projection of the stimuli on each plotted against their order in terms of their physical descriptions. Even with the third (psychological) dimension there were problems; this only provided three classes, asymmetric left, symmetric, assymetric right, and so only in the least restricted sense could it be considered a dimension in the same sense as the first two.

Torgerson goes on to argue, on the basis of these findings and of those of others, that similarity is not a unitary concept. He distinguishes between "... similarity as a basic, perhaps perceptual, relation between instances of a multidimensional attribute and similarity as a derivative, cognitive relation between stimuli varying on several attributes." The latter case he describes as more complex because of its susceptibility to attitude, attentional states and stimulus context, and hence the greater likelihood of the appearance of qualitative or class variables where it holds. However, he does not conclude that this invalidates the technique; rather he suggests, somewhat on the lines of the argument presented above, that the power of the MDS methods lies in their potential for making plain the structure underlying a set of similarity judgements, whatever the nature of the stimuli and hence of the type of similarity relation holding between them.

It/

It seems reasonable to conclude that this approach to the analysis of similarity data is both appropriate and powerful if care is taken in its use, in spite of the lack of strict criteria for the definition of the monotonic relation between similarity and distance and for the specification of the dimensionality of the final configuration. The most important point is that this should not be determined by one's expectations but by the data; if there is any doubt as to whether the solution obtained is in error, it can be checked either by re-running the data using different starting coordinates or by using a different method entirely. The terms in which the configuration is interpreted will depend to some extent on the nature of the stimulus set, but it is more likely to be profitable to consider it as representing clusters of similar objects grouped by class rather than as a representation of the axes of a "psychological space" along which the perceived members of the stimulus set are organised. This suggestion is certainly more congruent with the underlying principle, that degree of perceived similarity is reflected in degree of proximity in the final configuration.

The approaches to the analysis of similarity data so far considered have been essentially data reduction techniques. They are concerned with the development of an algorithm whereby a large set of measures can be transformed into a more compact and comprehensible form without too much distortion, and little attention has been paid to their theoretical basis. By contrast, a number of theoretical and experimental papers have been/

been produced which are directed at an analysis of MDS from the point of view of measurement theory (Beals, Krantz and Tversky, 1968: Tversky and Krantz, 1969a). The questions of the metric structure of similarity data and its representation by a spatial model are initially treated separately. In the former case a description is given of the properties which must be present in a set of dissimilarity (or similarity) data before it can be considered to embody a metric with additive segments, satisfying the triangle inequality. The (general) criticisms of the Shepard-Kruskal approach made by Beals et al are vitiated by their failure to appreciate the difference in the data set over which a metric is defined; Shepard and Kruskal deal with a set of distance measures based on an initially randomly-derived configuration, which is adjusted to conformity with the obtained similarity measures. These last are used only as a criterion of goodness of fit; no arithmetical operations are carried out on them. The final configuration does not represent the metric relations holding between the data points, but an approximation to these based on the best-fitting set of interpoint distances that can be achieved in terms of the monotonicity criterion. Thus the Beals-Krantz-Tversky formulation is dealing with a separate problem from the Shepard-Kruskal approach and therefore does not constitute a directly relevant critique of it.

The second part of their argument, the examination of the conditions under which a spatial representation of similarity data is appropriate, is treated at greater length. The models presented are based on generalisations of/

of the Minkowski, or power, metric, where the distance between two points $x = (x_1 \ldots\ldots x_n)$ and $y = (y_1 \ldots\ldots y_n)$ is given by:

$$d(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^r \right)^{1/r}$$

where r is equal to or greater than one. Where r is set equal to two this equation gives the standard Euclidean metric; where r equals one, it gives the so-called city block metric, where distance is not measured directly between two points but is taken as the sum of the distances between them on each dimension of the space.

Three assumptions are embodied in the general equation for the Minkowski metric. The first is that the distance between points is a function of "component-wise contributions." This is called decomposability. The second, embodying what is described as intradimensional subtractivity, says that each component-wise contribution is the absolute value of the scale difference, in other words that the distance between two points is a function of the difference between them on each dimension. The third assumption, interdimensional additivity, extends this by stating that the distance is a function of the sum of these differences. The first assumption gives the most general model. The latter two can be combined into what is known as the additive difference model, which is a special case of both its components. It should be emphasised that the equations embodying these assumptions do not require that the data to which they might be applied correspond to any metric. However, if both the metric and dimensional assumptions/

assumptions are to be satisfied it is shown that the only model which can accommodate both sets of constraints is the Minkowski metric.

The papers by Beals et al (1968) and Tversky and Krantz (1969) are devoted almost exclusively to the mathematical analysis of the various models presented. The details are not relevant here, but the point of view underlying their approach, and one of their extra-systematic assumptions, should be considered. The authors are concerned with MDS models not as data reduction techniques but as quantitative psychological theories. Their focus of interest, however, is on whether or not a particular set of data satisfied the assumption underlying their models, to the extent that it is suggested that the properties of subtractivity and additivity can be taken as defining the existence of psychological dimensions. It is also assumed (the extra-systematic assumption mentioned above) that a particular "factorial (i.e. dimensional) characterisation" of the stimuli is initially available. It is apparent, therefore, that the criterion for deciding whether or not a data set can be considered as embodying psychological dimensions is the congruence of the configuration representing the data with that representing the physical structure of the stimulus set. To adopt this point of view is to distort beyond the limits of normal usage the notion of a dimension, to overlook the possibility that a dimensional interpretation of the results is not necessarily the most appropriate, and that a class one may be necessary, and that similarity judgements may not reflect exactly the physical structure of the stimulus set.

These/

These comments, however, merely serve to suggest that this approach needs to be supplemented by theoretical considerations appropriate to the context when it is applied in a given experimental situation. The fact that it can be used to indicate where and to what extent perceived structure deviates from physical structure makes it a powerful normative model in the sense used earlier. Requiring a preliminary description of the structure of the stimulus set in terms of the model means that highly specific predictions can be made, and that the reasons for any failures in prediction that may occur can be examined in the light of the discrepancies, thus providing some insight into the subject's processing of the stimuli. It is this tightening of experimental control that gives this model the advantage over that of Shepard and Kruskal, particularly in view of the arbitrary decision criteria associated with the latter.

Before leaving the topic of methods of analysis, a brief account should be given of an alternative approach. This is the hierarchical clustering scheme (HCS) presented by Johnson (1967). The object here is to represent the N(N-1)/2 similarity measures in terms of a hierarchy ranging from the initial level where each object in the set from which the data has been derived is taken as forming a cluster including only itself to the ultimate level where the whole set forms a single cluster. The method provides a straightforward alternative to MDS techniques for reducing and clarifying similarity data.

The procedure for deriving an HCS is simple. The second level of clustering/

clustering (the first, as already stated, is simply that where each object constitutes a cluster) is defined as consisting of the two objects which are judged to be more similar, in terms of whatever experimental measure has been used, than any other pair. The value of the similarity relation between them is taken as the value of the cluster. The data matrix is then rewritten with one less row and column, the two subsets of similarity measures associated with the objects forming the cluster being merged into one. This is done by retaining the larger value of those between the clustered objects and all others; thus if A and B form a cluster, and the similarity measure between A and C is 25 and B and C 20, that between the cluster A/B and C is taken to be 25. This procedure is repeated until the final level is reached. The use of maximum values in defining the clusters ensures that monotone invariant results are obtained. It is clear that some information is lost in comparison with that obtainable from MDS techniques, and that the relations between all stimuli are not brought out to the same extent. Nevertheless the HCS represents a simple way of bringing out the basic partitions of the stimulus set, as will be seen.

Having discussed at some length the general apparatus available for the study of similarity judgements, we now turn to situations where it has been applied with a view to seeing what factors have been shown to influence performance. Three topics will be taken up here; the nature/

nature of the metric underlying the data, and, related to this, the different ways in which analysable and unanalysable stimuli are processed; and whether it is more appropriate to study experimental findings on a group or individual basis.

There is now a considerable body of evidence (Shepard 1964; Torgerson 1965; Garner 1970) that analysable and unanalysable, or homogeneous, stimuli require different metrics for the spatial representation of similarity judgements based on them. In the latter case, for example, colour patches, the data correspond quite closely to the Euclidean metric, where distance is measured directly between two points. However, where the stimulus figures are composed of distinct attributes this no longer holds, and the city-block metric, where distance between two points is measured by taking the mean of the absolute differences between the points on each dimension, has been shown to give a more accurate account of the similarity relations. The inference based on this distinction is that where the measures of similarity conform to the city-block metric, subjects are analysing stimuli into their component parts and basing their responses on a direct combination of the differences. When the Euclidean metric is more appropriate, however, it is suggested that overall impressions of similarity, not involving such a decomposition of the stimuli, are being formed.

The nature of the underlying metric can be assessed either directly or indirectly. The Shepard-Kruskal MDS model is an example of the indirect approach; if the interpoint distances derived from one or the other/

other provide a better fit to the similarity measures for a configuration in a given number of dimensions, then it can be assumed that the data correspond more closely to that metric. The direct approach involves the concept of an isosimilarity contour, that is, the locus of all stimuli with a given level of similarity to the standard (Shepard 1964; see also Beals et al, 1968). With the Euclidean metric the shape of this locus, or line joining points of equivalent similarity, is elliptical, but if the city-block metric holds it becomes a diamond. An example of how data can be manipulated to bring out information bearing on this is provided in the Shepard paper.

Garner (1970) gives an extended analysis of the analysable/unanalysable distinction in the context of a discussion of the nature of the stimulus. He talks of integrality rather than analysability, contrasting it with separability. Two dimensions are said to be integral "... if in order for a level on one dimension to be realised, there must be a dimensional level specified for the other." A visual stimulus has a brightness, hue, saturation, size and shape; thus any pair of these dimensions are integral. Two dimensions are separable, on the other hand, if they can be present independently of each other. Garner describes these as idealised, or limiting, definitions, the reason for this being "... the versatility and adaptability of the organism..." He reviews much evidence which suggests that stimulus dimensions can be redefined by subjects, and comes to a conclusion close to the point of view expressed throughout this chapter, that only by paying more attention firstly to stimuli and secondly how they are perceived can soundly-based conclusions/

conclusions about information processing be reached.

The redefinition of stimuli brings us to our final point, the question of group or individual analysis of results. Both Shepard (1964) and Torgerson (1965) found that distinct strategies were being used by subgroups of their subjects, in the former case to the extent that if the groups had not been studied separately it would have appeared from the data that a metric representation of it was simply not possible. Thus although an analysis of each subject's performance may be redundant, it is nevertheless necessary to be aware of the possibility of producing misleading results by assuming that all subjects are treating all stimuli in an equivalent fashion.

The concept of similarity, in the context of information processing tasks in general and form discrimination in particular, is clearly not straightforward or easy to handle. Its complexity arises from its vagueness, and the wide range of objects, situations and processes where it is invoked. This in turn imposes severe constraints upon any investigation involving aspects of similarity, since otherwise it is not possible to come to firm conclusions on the processes involved. The most important thing is to provide a normative model in the sense of a rigorous, unambiguous and exclusive description of the stimulus set used. This must be accompanied by a method of analysis capable of isolating the discrepancies, if any, between the physical and perceived structures of the stimulus set. Such methods are available, in particular the Beals-Krantz-Tversky approach. The optimal stimulus specification would seem to be in terms of discrete attributes, varying/

varying in level, forming what Garner describes as a complete orthogonal set, that is one where no dimensions of variation are correlated. Such a description fulfils the criteria laid down above, and at the same time can be made sufficiently simple to reduce the information overload on the experimenter. The form psychophysics approach, although rigorous in its own way, not only does not measure up to these requirements but builds in potentially destructive factors by its acceptance of such concepts as sampling variability and residual error.

No mention has so far been made of the empirical questions raised at the beginning of this chapter. They are raised as part of the introduction to the next set of experiments, where the theoretical and methodological issues raised here will be tested.

Chapter 4

The detection and use of stimulus information in form discrimination tasks

The work to be described here consists of a number of cross-sectional studies, by contrast with the longitudinal approach of Experiment I. In general, we wish to examine in more detail certain aspects of perceptual functioning in the context of a test of the ideas presented in the previous chapter. As in the first series of experiments, both orientation and form discrimination are studied; the objective is a more precise account of the ability of pre-school, and slightly older, children to detect and reproduce the structural information present in a stimulus set. Two more specific points will also be considered; the extent to which response patterns differ when matching-from-sample and pair comparison designs are used, and the occurrence of asymmetry of response in relation to the nature of the attributes comprising the stimulus set. In the former case it is expected that the child's ability to reproduce the structure of the set will be worse in the pair comparison situation, since the total set is not available for inspection at any time. This follows from the findings of Over and Over (1967) and Bryant (1969), that performance is impaired in young children when successive rather than simultaneous discriminations are required of them, and the consequent inference that their memory for stimulus features is poor.

In discussing the asymmetries found in Series II and III of Experiment I, Woodworth's/

Woodworth's schema-plus-correction was invoked; it was suggested that a figure rotated from its "normal" orientation is seen as being in some sense more complex than is the non-rotated version, and that as a result the latter is more likely to be matched to the rotated figure than vice versa. The work of Handel and Garner (1966), on the relationship between pattern goodness and pattern association, is relevant here. They found that patterns with a low "goodness" rating, and hence from a larger subset, had many different associates, whereas those with higher ratings had very few. Thus a poor pattern can be associated to one with a higher "goodness" rating, but the converse does not hold. The analogue of this situation in a form discrimination task is that the unrotated figure represents the good pattern and the rotated one the poor pattern, and hence that the rotated figure is more often matched to the former than vice versa. We might anticipate, therefore, that where orientation is one of the attributes differentiating figures, asymmetries will be found to a greater extent than where it is not so used.

The normative model to which the results of these studies are referred is based for two of the three stimulus sets on that used in an experiment by Tversky and Krantz (1969b). This was designed as a test of their inter-dimensional additivity hypothesis, in such a way that departures from expectations could be accurately pinpointed. Their stimulus set comprised eight schematic faces, composed of the three attributes of overall shape, eyes and mouth; each of these had two levels or values (see Fig. VI).

According/

According to their hypothesis the judged dissimilarity of any pair of these figures is an increasing function of the sum of the differences between each figure on each attribute. Three predictions are based on this; that dissimilarity is constant over all stimulus pairs differing by one, two or three attributes within the same class; that the dissimilarity increases if new differences are added, that is, the dissimilarity of a stimulus pair differing in two respects is less than that of a pair differing in three and greater than that of a pair differing in one; and that the ordering of pairs differing in two respects is determined by the ordering of pairs differing in one. Thus if the dissimilartty of figures differing in eyes is less than that of figures differing in mouth, then the dissimilarity of figures differing in eyes and overall shape is less than that for figures differing in mouth and overall shape. The predictions were fulfilled in an experiment requiring numerically-expressed judgements of similarity. The data for each subject were analysed individually; with six subjects, four out of the six possible orderings of attribute classes were obtained, which would have produced a very muddled picture if the results had been combined. With this approach, then, quite specific predictions can be made of the ordering of stimulus pairs in terms of the judged similarity between them. A similar, although less rigorous analysis is applicable to that section of the experiment where orientation is the sole distinguishing feature, as will be shown.

A/

A formal model is thus available which provides the kind of stimulus specification that is required and at the same time allows a number of different methods of analysis to be applied to the data. Before going on to describe the experiments proper, however, some attention must be given to the empirical questions raised in Chapter 3, particularly those concerned with the nature of the decision rule used in matching stimulus figures and the extent to which differences can be detected. The third of these questions, the adequacy with which discriminations are being made, can of course be answered only in terms of the results of any given experiment. It is clear from the evidence provided in Chapter 1 and the findings quoted in Chapter 2, that children around five years of age tend to equate partial and complete identity. Thus their decision rule, which embodies their notion of what is meant by "the same as", will lead them to call two stimulus figures "the same" if there is some degree of perceived similarity between them. Series III of Experiment I also suggested that there are quite tight restrictions on what constitutes perceived similarity, in that responses indicating this were very rarely made when pairs of figures differed in more than one respect.

However, this raises the second question, how well children can detect points of difference. The studies of Vurpillot (1968) and Zaporozhets (1964) show that children of the same age as the subjects used here have poorly developed strategies for the scanning and comparison of figures. The inference is, therefore, that not all differences are picked up. This leads into a circular problem, in experiments of this nature; does the fact that the decision criterion for identity corresponds to partial equivalence stem from the/

the child's incomplete processing of the stimuli, so that he "thinks" that two figures are identical because he has not succeeded in identifying any difference between them, or is the scanning process under the control of the decision rule so that as soon as some degree of equivalence has been established between the figures, the matching criterion is satisfied and comparison is discontinued.+ These alternatives suggest different outcomes in terms of response pattern. Assume that in the latter case, with stimulus figures defined by three attributes, the recognition that any two attributes have the same value in both figures of a pair meets the decision rule criteria. Thus all pairs of stimuli differing in one attribute would be responded to as "same", as would all physically identical ones; the remainder would be successfully differentiated. If, however, it is the child's ability to detect features of difference that is limiting his performance, then we would expect certain of them to be consistently over-looked or ignored; certain attribute differences will be discriminable while others will not. The response pattern here will show those pairs of figures differing by a non-discriminable attribute not being distinguished, whereas the remainder will be. If the Tversky and Krantz ordering hypothesis is correct, responses to figures differing in two attributes will show a comparable pattern; where the non-discriminated attribute is one of the two differentiating stimulus pairs, more confusion errors will be made than where it has the same value on both stimuli. At the same time, because of the presence of a distinctive attribute in the former case, fewer errors would be expected than where only the one, non-detected, attribute marks the difference.

If/

+ On the assumption of random scanning

If this argument is correct, we have a prediction which can be tested out in terms of the dimensional representation of the perceived structure of the stimulus set. If all differences up to a certain limit are ignored, as in the first alternative given above, then the configuration should show all those figures differing by the value defining this limit to be clustered together; if one particular difference, by virtue of its being less discriminable than the others, is consistently overlooked, then the stimulus figures can be expected to cluster in pairs defined by the other two attributes they have in common.

There seems little point in making any suggestions as to the nature of the metric underlying the similarity measures. Of the two stimulus sets to be used here one is composed of separable attributes and one of integral; but since we know that children differ from adults in their processing of material of this nature, it seems reasonable to adopt a wait-and-see policy. Indeed the child's capacity for redefining stimulus sets may well be so great that questions of the type of metric to which his judgements correspond become redundant.

The experimental work is divided into three parts, the same subjects and material being used in all three. A general description of these is given, followed by more specific accounts of the hypotheses and findings for each part.

Subjects:

Five groups of subjects were used, drawn from different nursery and primary/

primary schools. Their composition is shown in Table VIII. There were 32 girls and 25 boys, giving a total of 57.

Table VIII

Group	N	Mean Age	Age Range
I	17	6-1	5-3/7-6
II	9	5-4	4-10/5-8
III	8	5-4	5-1/5-7
IV	13	4-10	4-3/5-7
V	10	4-2	3-10/4-8

Subject Groups

Results will be presented for the groups singly and together; the reason for not amalgamating them from the beginning is that there were considerable differences between them not only in age but also in social and cultural background, which might be expected to influence the results. Thus Group I was drawn from a primary school in a predominantly working-class urban district, while Groups IV and V came from well-equipped city nursery schools and Groups II and III consisted of children from the kindergarten class of a rural primary school.

Material:

The stimulus sets used here are shown in Figs. V and VI. The orientation discrimination/

discrimination problems were designed to provide as much variation as possible so that the form of analysis used in Experiment I could be extended. The other two are defined by orthogonal attributes, that is no attribute or level of attribute is correlated with any other. Each member of each of these sets is described in terms common to all. The schematic faces, taken from Tversky and Krantz (1969b), are composed of three attributes each of which can take one of two values; overall shape (elongated either horizontally or vertically), eyes (filled or blank), mouth (straight or curved). These attributes are separable according to Garner's (1970) definition. The circle figures of the third set are defined by two attributes, each with three values; orientation (upright, rotated 45° right, rotated 90° right) and angle subtended by gap (20°, 90°, 180°). The order of presentation of the three sets was varied over groups.

In view of the wide age range of the subjects used here, there will be a number who fall in each of the three performance categories described in Experiment I (Table V). Thus in order to maximise the number of responses available for the kind of analysis proposed here, subjects were pressed to produce more than one for each discrimination problem. In all the studies reported here, if only one comparison figure was matched to the standard in response to the question "Can you see one up here that's just the same as this one", the subject was asked "Are there any other ones just the same as this one", "this one" being the standard. Where more than one comparison figure was given as a match following the first question, the supplementary one was not asked.

Experiment II: Orientation discrimination

The purpose of this experiment is to examine in greater detail the findings of Experiment I, that global features of similarity were detected and used in orientation discrimination and that the orientation of the standard figure in relation to the members of the comparison set was more likely to affect responses than the absolute value of the difference between the standard and any one comparison figure. The stimulus arrays used were designed to show this effect as clearly as possible. In view of the strong position preference found in the previous study for choosing figures in position 3 and 4 in a six-card linear array, comparison figures identical to the standard were not placed in these positions in any of the problems. The method of presentation and the instructions were identical with those used in Series I of Experiment I, with the addition of the supplementary question where only one match was given.

Results:

The results of Experiment I suggested that the systematic use of global aspects of similarity in an orientation task similar to that used here was accompanied by an increase in number of multiple responses (MR), that is responses where more than one comparison figure was matched to the standard, and by an increase in number of correct responses. These increases were in relation to the pattern characteristic of what was described on Stage A, where few MR were given and number of correct responses was low;/

low; this was held to represent the lowest level of ability, where choice of match to standard was essentially random. Since the focus of this experiment is on regularities in the use of stimulus information, it is necessary to establish first whether or not the subjects used here have passed this initial stage. Clearly number of MR cannot be used as a criterion, since the study was designed to force these as much as possible. However, the opportunity was still available for them to be given spontaneously; that is, subjects were free to make more than one match following either the initial instruction or the supplementary question. These spontaneous MR will be used here as an index of level of performance.

It should be pointed out that what follows is based on group rather than individual data. As will be seen, there are insufficient constraints in the stimulus material used here to permit the kind of unique specification that was described in the previous chapter; thus a discussion of response probabilities rather than a classification of subjects is more appropriate. In addition the identification of change over time, the focus of Experiment I, is not at issue here; but nothing is lost by working with group measures since the kind of pattern we are interested in here can be demonstrated quite clearly in these terms.

Before turning to the analysis of use of stimulus information the more basic features of the results should be considered. The data is given in Table IX. From the figures in the first two columns, it is obvious that ability/

ability to make correct discrimination is a function of chronological age; the oldest group (I) give proportionately more correct results than do Groups II, III and IV, who in turn are more successful than the youngest, Group V. The overall drop in proportion of total responses correct compared with first responses merely indicates that when pressed to give more than one response in a situation where only one is correct, subjects will make more errors. Nevertheless, the orderings in terms of both first and total responses are remarkably consistent. The third column, however, shows that Group III, one of the middle-range groups, make fewer omission errors than do their peers.

Table IX

Group	First responses correct		Total responses correct		Total responses correct as proportion of max. poss. correct responses	Total responses	Average spont. MR subject
I	97	(72)	125	(49)	92	252	1.7
II	30	(43)	50	(34)	69	145	3.4
III	27	(43)	59	(34)	92	174	7.2
IV	43	(41)	73	(34)	70	214	1.9
V	22	(31)	38	(25)	49	152	3.3

Total responses and correct responses for each subject group: Bracketed figures in first two columns show correct responses as proportions of total, decimal points omitted. See text for further explanation of third column.

The figures here represent the total number of correct responses (second column) as a proportion of the possible total of correct responses; for each group this latter figure is given by multiplying the possible number of correct responses for each subject, which is eight, by the number of subjects in the group. Thus the greater the values in this column, the smaller the number of occasions in which the subjects of a given group failed to make the correct match to the standard. Group III's success here would appear to be due to the total number of responses they made. It can be seen from the last column of Table IX that they made more spontaneous MR than any other group; but they also made a greater number of choices, the average for each subject being 22 compared with those for the other groups of between 15 and 17. The fact that Group III performs at the same level as Groups II and IV in terms of proportion of first responses correct indicates that they are not capable of better discrimination as such, simply that by giving more responses they have increased their chances of hitting on the correct ones.

The criteria for inclusion in category A, representing the lowest level of performance, were given in Chapter 2 as less than 25% MR and less than 25% correct responses. Here 25% MR is equivalent to 2. In these terms, therefore, none of these groups can be classified at this level, none of them fulfilling both requirements (final column and bracketed figures, second column, Table IX). Thus it can be inferred that although there is doubtless some noise in the results, they are essentially non-random. It has already been/

been noted that the proportion of total correct responses is artificially lowered here; the fact that even under these circumstances the performance of all groups can be classified at a higher level than category A increases the force of the conclusion that their responses are systematically based.

The extent to which responses consistently reflect stimulus features can be assessed in the following way. Under the circumstances of this experiment, perfect performance would require that all subjects in a given group give the correct match to the standard as their first response, and reject the suggestion that any other comparison figures were identical to it. One or two in fact did this. The next level down from this would involve no errors being made on first response, but a number of additional responses being offered. Since we are assuming, for the moment, a steady departure from complete accuracy, these additional responses can be expected to be to the comparison figure(s) perceived as most similar to the standard. Thus in terms of group results, as performance drops away from the optimum level, we would expect frequency of response to reflect degree of similarity between the standard and comparison figures; and by examining the order of the members of the comparison set, the order being defined in terms of the frequency with which they are chosen as a match to the standard, we can see directly what the degree of perceived similarity is between them and the standard. As far as first responses are concerned, it would be expected that where these were not all correct, the alternatives chosen/

chosen would reflect the order of perceived similarity between figures in the same way; and to the extent that responses reflect a systematic but not perfectly accurate use of stimulus information, we would expect the order of first responses to be identical to that of total responses.

Table X gives this information for all groups combined. It shows the order of frequency of response to each comparison figure for each problem, for both total and first responses (the latter in right-hand columns), the numbers in the body of the table being card positions. Thus for problem 1, the largest proportion of the total number of responses is made to the figure in position 1, which is the correct response; the next largest proportion of responses is to the figure in position 3, the next to that in position 5, and the next to those in positions 4 and 6 jointly.

Table X

Problem 1		2		3		4		5		6		7		8	
1*	3	5*	5	6*	6	2*	2	5*	5	1*	1	6*	6	2*	2
3	1	1	1	2	2	3	3	2,3	3	5	5	3	3	5	3
5		3		3	3	6	6	6	2	3	3	1		3	5
4,6		6		1		5	1			6	4			6	6
						4				4					
						1				2					
77	93	72	87	73	91	70	82	91	89	85	83	87	88	90	87

Similarity ordering of comparison sets in terms of frequency of response, totalled over all subject groups, for all problems. See text for full description.

Thus/

Thus the further down the column a figure is, the less similar to the standard it can be taken to be. The asterisks indicate the correct matches to the standard; only in one instance, first responses to problem I, is this not the figure with the highest frequency of response. Numbers on the same line indicate equal frequency of response to the comparison figures in those positions, for example, 4 and 6, problem I. Where comparison figures do not appear in a column, this indicates that less than 5% of the number of responses was made to it. The numbers at the foot of each column indicate the percentage of total number of responses made to the stimulus figures displayed in the positions given above the line in each column; for example, taking first responses to problem I, 93% of these were made to positions 1 and 3, the remaining 7% being distributed over the other four positions.

In terms of the assumptions laid down previously, we are looking for congruence between the orders for first and total choices for each problem. In three cases, problems 1, 6 and 8, this does not hold, but in each the deviation consists of the reversal of two positions in one order relative to the other. Problem 5 at first sight may appear to represent another violation of this requirement, but the fact that figures in positions 2 and 3 are unordered with respect to each other in terms of total number of responses allows the two orders to be taken as equivalent. Thus even where results from groups of mixed ability are combined, considerable regularity of response is demonstrated.

The results presented in Table IX indicate that the older the subjects

in/

in a group, the better is that group's performance. A similar, although not identical, impression is obtained from the breakdown of Table X into the results for each group. These are summarised in Table XI.

Table XI

Group	1	2
I	1	0
II	5	4
III	2	3
IV	5	4
V	5	9

See text for explanation.

Column I gives the number of violations of order consistency over the eight problems, while column 2 shows the number of times the correct match to the standard does not show the highest frequency of response. The possible maximum here is 16, there being two orders for each problem. Group I are clearly the most accurate, in that correct responses are in all problems made more frequently than incorrect cnes, and also the most consistent. Of the three middle-range groups, Group III is marginally superior to Groups II and IV in both respects, confirming the notion suggested by their low number of omission errors that in spite of the number of responses they gave, greater than that of any other group, they were in fact operating under greater systematic constraints than either of their peer groups inasmuch as/

as the distribution of their responses is more restricted. Group V can be seen to be less accurate than any of the others, but no less consistent than two of the older groups.

So far what has been said has been intended to demonstrate that the results are not random and so that it is legitimate to take frequency of response to a member of the comparison set as an indix of the perceived similarity between it and the standard. We now turn to an examination of the stimulus figures in order to see what kinds of features responses are based on. This will be done for each problem individually.

First responses to problem I (Table X) show quite clearly that the two horizontally-oriented figures with the opening of the "hook" upwards are the most obvious first choices. However, the incorrect response is given more often than the correct one. This raises the question of a position bias, similar to that found in Experiment I; this possibility will be considered in more detail at a later stage. Be this as it may, no other figures are initially considered to be the same as the standard. Moving to the order of total responses, the correct choice can be seen to be more frequently made. The same two figures still occupy the first two places in the order, and the third most frequent response is to the only other horizontal figure in the array. The two oblique figures complete the distribution. What was described as an "obvious mis-match" in Experiment I, here represented by the vertically-oriented figure in position 2, is not confused with the standard at/

at all. Here, therefore, the order of similarity is horizontal figures, obliques and vertical figures, with the last mentioned only for the sake of completeness. In addition to the possible influence of a position preference effect, the role of distinctive features should be considered. In both oblique figures the "hook" is on the same, left, side of the card as it is on the standard figure. This question also will be raised again at a later point.

In all remaining problems the correct match to the standard is the most frequently chosen; comments will therefore be restricted to the order of similarity. In problem 2 the most similar non-identical figure to the standard is the 90° rotation in position 1. The two other oblique figures in the array are next, while the non-oblique ones are virtually never chosen. The vertical figures in problem 3, corresponding to the orientation of the standard, occupy the first three places in both orders. Following the correct match, the next similar figure is the one which represents a rotation of the standard about its vertical axis in the third dimension, so that the larger part of the figure is still at the top. The third vertical figure does not preserve this relation, and is perceived as less similar to the standard. Here the horizontally-oriented figure does not appear in the order.

Problem 4 indicates the influence of distinctive features. All members of the comparison set are obliquely oriented, but the two which have the double lines at the bottom end in contrast to the top end, as on the/

the standard, are seen as least similar to it, with the figure in the same orientation as the standard being last in the similarity order. The shape chosen for problem 5 can be described in Wohlwill and Winer's term as having "low directionality"; nevertheless, here again there is consistency, with the two vertical figures not being described to any extent as similar to the standard. The frequency of choice of the figure in position 3, and its consequent appearance in one order above the other horizontal figure and in the other as tied with it may well be a position effect. In problem 6 the total response order is nicely consistent with what one would hope to find; the oblique figure in the same orientation as the standard is seen as more similar to it than the other two obliques in the opposite orientation, while the vertical and horizontal figures are least similar. This does not appear in the first response order, however, where the vertical figure appears above one of the obliques. The two vertical figures in problem 8 account for by far the largest proportion of responses, with one of the oblique figures, rotated 45° from the orientation of the standard, being chosen a few times only.

Finally, the total response order for problem 8, like that for problem 6, is very neat; it is headed by the two bottom-heavy horizontally-oriented figures, followed by the two top-heavy ones; the obliques are rarely chosen.

Before going on to discuss these findings, some attention should be given to the question of the presence of a position preference or bias effect.

Table/

Table XII shows the distribution of responses over positions in the array.

<u>Table XII</u>

Group	Card Position 1		2		3		4		5		6	
I	19	21	17	22	16	11	4	7	20	24	23	23
II	12	15	15	15	26	26	11	12	19	19	16	15
III	16	15	17	15	20	27	3	3	21	19	22	20
IV	17	16	15	18	22	27	5	4	21	22	20	14
V	15	18	22	19	26	35	12	6	16	10	19	12

Proportion of responses to each position in the comparison array (decimal points omitted), taken over all problems for each subject group. Left-hand columns, total responses: right-hand columns, first responses.

If performance were completely accurate, 25% of first responses would be made to figures in positions 1, 2, 5 and 6, since the correct choice appeared twice at each of these positions over the eight problems, and there would be no responses to positions 3 and 4. As can be seen, Group I approximates most closely to this. The remaining groups give a greater proportion of their responses to position 3, and this increases as the mean age of the group decreases; at the same time responses to position 4 are, with the exception of Group II, uniformly low, and even Group II make less than half the number of responses to position 4 that they give to position 3. The/

The fact that responses to position 3 are proportionately greater in the younger, and hence, as has been seen, less accurate and less systematic, groups suggest that there is some bias in the overall results, but the low level of response to position 4 also indicates that the distributions do not reflect this alone.

Conclusions:

As would be expected, accuracy of response in terms of both correct responses and omission errors increases with age. However, we have also seen that responses are more systematic with older subjects, in that they are more likely to detect the correct match to the standard initially and to show consistency in their non-correct judgements. The relationship between extent to which responses reflect gross features of similarity, number of correct responses, and number of multiple responses, brought out in the longitudinal study reported in Chapter 2, has been shown to hold under the modified circumstances of this experiment; and it has also been shown that the better the performance in terms of the classification scheme described earlier, the less likely it is to be based on factors other than stimulus information.

The analysis of the kinds of confusion most likely to appear in problems of this kind confirms that it is not the absolute nature of the difference between the standard and the comparison figures which is relevant, but rather the relationship between them; in problem 1, for instance, a

90°/

90° rotation of the standard was virtually never confused with it, whereas this was the most common error in problem 2. The similarity orderings suggest that it is possible for children of this age to respond differentially to obliquely-oriented figures in different orientations, but only when asymmetrical figures are presented so that the contrast between the two orientations becomes apparent; otherwise they will respond to oblique figures as such, regardless of orientation.

However, these results also suggest that orientation, at least with the stimulus figures used here, is not the sole determinant of response. The presence of a feature in a figure which can be used to differentiate between its presentations, for example, the "hook" or problem 1 and the "prongs" of problem 4, can also provide information which may be used. And the data presented in Table XII suggest that for younger children there is an interaction between position in the array and degree of similarity to the standard. Thus it could be suggested that the frequency with which figures in position 3 were responded to reflects the fact that this was the first position looked at, and that the figures at that position were more often than not at least to some extent similar to the standard, more so certainly than was the case with position 4. However, a record of eye-movements would be required for a proper test of this hypothesis.

The results can therefore be taken as indicating that relational features in the stimulus set are detected in an orientation discrimination task, that these are reflected in subject's responses, but that the features used/

used and the extent to which irrelevant (to the experimenter) factors determine responses depend very much on the nature of the figures used as stimuli. Garner's comment about the subject's ability to redefine the stimulus is perhaps the most relevant one for this kind of task, where whatever information is available may be used and even some that is not.

<u>Experiment III: Form discrimination</u>

This study explores the ability of children to detect and use, in a matching task, information defining stimulus sets the members of which vary in respect of more than one attribute. With the two sets used here (Fig. VI), the relationship of each stimulus figure to every other can be completely specified in terms of the attribute structure of the set. The formal description of the structure, together with the ordering hypotheses mentioned earlier in this chapter, provide the normative model in terms of which subjects' performance is assessed. There are three classes of difference between figures for the schematic faces (set 1); (p. 81) since these vary in three attributes, each of which can take one of two levels, the differences can be in one, two, or three attributes. With the circle figures (set 2), however, there are five classes, since the figures are composed of two attributes each of which can take three values. The classes of distinct figures are; one attribute constant, the other differing by one or by two levels; both attributes differing, by one or by two levels; and one attribute differing by one level, the other by two.

Experimental procedure here was identical to that used in Series II and III, Experiment I, a duplicate of each figure in the set being placed in the centre of the bottom row (set 1) or below the array (set 2). Where every member of the set is compared with the whole in this way, conclusions on the processes involved are based on subjects' perception of the structure within the set, assessed by a series of converging operations, rather than on generalisations/

generalisations across results from different sets as in the previous experiment. By speaking of "converging operations" here it is intended to emphasise the fact that subjects have the opportunity to define equivalent subsets in several different ways. For example, suppose that a subject indicates that figures A, B and C in the array are the same as the standard figure A. If with B as standard he makes the same three choices, then it can be inferred with a greater degree of certainty that these three figures are not differentiated as far as he is concerned than would be the case if only one set of results were obtained. Thus the investigation of the detection of structural relations is more limited, but also more rigorous.

Table XIII

Group	First responses correct		Total responses correct		Total resp. corr. as prop. of max. poss. corr. rep.	Total responses	Average spont. MR/subject
I	82	(60)	128	(40)	94	319	3.8
II	38	(53)	57	(42)	79	135	2.7
III	25	(39)	50	(31)	78	164	5.8
IV	48	(46)	86	(33)	83	262	3.5
V	19	(24)	51	(27)	64	191	5.0

Correct responses, etc. for all groups in set 1 (faces. Format as for Table IX

Results:/

Results:

The findings from the two stimulus sets will be discussed separately, and then brought together in the analysis of the asymmetry data. Table XIII gives the basic results obtained from the schematic faces. On the same criteria as used in Experiment 2, it is clear that responses can be taken to be non-random, since no group show a level of performance corresponding to that of category A. The correlation between age and accuracy is still apparent here, although it is less definite than in the previous study. The proportions of total correct responses are low, but this can be attributed as much to the large number of responses made by all groups as to any other factor. The third column, indicating omission errors, gives a more tidy picture; the age trend is more clearly defined, and the overall accuracy of matching can be seen to be higher than consideration of the first two columns only would suggest.

The focus of interest, however, is less on how accurately children can discriminate these figures than on how their incorrect matching responses are distributed. Table XIV shows this in two ways:

Table XIV

Group	Degree of Difference					
	1 attribute		2 attributes		3 attributes	
I	73	(79)	25	(21)	2	(0)
II	81	(88)	19	(12)	0	(0)
III	71	(90)	28	(10)	1	(0)
IV	67	(73)	30	(25)	3	(2)
V	62	(57)	34	(41)	4	(2)

Left-hand columns, proportion of wrong responses made to comparison figures differing from standard in 1, 2 or 3 attributes. Right-hand columns proportion of wrong first responses involving each degree of difference.

in terms of the distribution of erroneous responses to figures differing from the standard as a proportion of the total number of responses, and of wrong first responses as a proportion of the possible total of first responses. To invert one of the Tversky-Krantz ordering hypotheses, it is quite clear that as the similarity between pairs of stimuli, in terms of attributes in common, increases, the number of confusions increases also. Taking Group I as an example, 60% of their first responses were correct (Table XIII); almost 80% of their wrong first responses involved judging a figure differing from the standard by only one attribute to be the same as it. With the possible exception of Group V, in all cases a much smaller number of errors involve figures differing by more than one attribute from the standard; even for Group V the difference is in the right direction.

Owing/

Owing to lack of computer facilities, it was possible to carry out an MDS analysis on only two sets of data, for Groups I and IV; and only that for Group I used both the Euclidean and City block metrics. The same program was used as in Experiment I. Virtually no difference was found in the adequacy with which each metric represented the data, in terms of the departure from monotonicity of the function relating distance in the configuration to judged similarity of the stimuli; Table XV gives the relevant information.

Table XV

Metric	Dimensions		
	2	3	4
Euclidean	.049	.030	.003
City block	.050	.030	.004

Values of Delta (index of departure from monotonicity) for Group I data using different metrics.

In both cases a four-dimensional solution gives the best fit to the data, but there are no real differences between the values of Delta. This suggests that either a mixed strategy is being used by the subjects, or that different approaches to the stimuli are being taken by subgroups, some of whom will be treating them as analysable and so giving results more/

more appropriately handled by the city block metric, and others responding in more global terms, forming overall impressions, which would produce a similarity ordering more congruent with the Euclidean distance measure.

The three-dimensional configuration for both analyses of the Group I data are shown in Fig. VII. This means of representation was chosen rather than that provided by two two-dimensional plots in order to bring out groupings of figures in contrast to their projections along dimensions. With both metrics, the groupings shown in three dimensions are preserved in the fourth; the latter does not introduce a greater separation between members of a group than is apparent in the plot given. The basic division of the stimulus set is between figures 1, 3, 5 and 8, which fall in the top half of the configuration, and 2, 4, 6 and 7, in the bottom half. This corresponds to the distinction between figures with a vertically-elongated overall shape and a horizontal one, respectively. Within this division the figures are separated in terms of whether or not the eyes are the same. This is particularly apparent for the Euclidean configuration, where 1 and 5, 3 and 8, 2 and 4, and 6 and 7 form quite distinct clusters. Thus it can be inferred that figures differing only in the shape of the mouth are more likely to be confused than any others, and that values of the other two attributes are more frequently detected and used.

The way the figures are clustered can also be made more obvious by using/

using the hierarchical clustering scheme described in the previous chapter. Fig. VIIIshows the hierarchies for all five groups. With the sole exception of that for Group IV these all show the first four clusters to be of the pairs of figures differing only in the shape of the mouth, again demonstrating that this is the attribute most likely to be ignored when describing the similarity between figures. Clusters beyond this are of two distinct kinds; that for Group IV brings together the two clusters which have the same type of eyes, showing that this is a more powerful determinant of similarity than overall shape, while Groups II, III and V bring together the sub-clusters composed of figure with the same overall shape; for these subjects, then, this is a more potent criterion of similarity than the eyes. Group I show a more mixed pattern. The fifth cluster brings together two of the lower-level clusters with the same type of eyes, but the sixth does not unite the other two; here the fifth is combined with the first, to form a cluster with two figures with the same overall shape as two of the four others but otherwise distinct from them. This loss of regularity and the resulting complexity of the similarity relations among the whole set of stimulus pairs was reflected in the MDS analysis, in the number of dimensions required to represent the data adequately.

The results described so far indicate quite clearly that where matching responses are incorrect they are nevertheless systematic, in that frequency of errors decreases considerably where the difference between figures/

figures is marked by more than one attribute. We now turn to the second ordering hypothesis mentioned earlier, that the order of similarity for pairs of figures differing in two attributes is determined by the order of pairs differing in one. This can be done for both group and individual data, in a similar way. The order of similarity is given by the number of confusion errors involving figures differing by a given amount; the more errors there are, the more similar the figures are seen to be. Table XVI shows the six possible orders which can be derived from this stimulus material; the "Total" column, giving the frequency of occurrence of each order, indicates that they are not all equally present, and that by far the greatest number of confusion errors involve figures differing in mouth. The fact that the sum of the entries here is less than the number of subjects is a reflection of the lack of errors produced by some of them.

Table XVI

1 - attribute difference	2 - attribute difference	Total
M: E: S	ME: MS: ES	20
M: S: E	MS: ME: ES	12
E: S: M	ES: ME: MS	5
E: M: S	ME: ES: MS	4
S: E: M	ES: MS: ME	4
S: M: E	MS: ES: MS	3

List of all possible orders of stimuli differing on one and two attributes. ":" stands for "are more similar than;" e.g. in row 1, figures differing in M(outh) only are more similar than figures differing in E(yes) only, etc. "Total" column gives no. of subjects producing responses wholly or partially corr. to each order.

Where responses are consistent with the ordering hypothesis, the order defined by the confusion errors for each subject will correspond to one of these. For example, if a subject confuses figures which differ in mouth only more frequently than figures differing in eyes only, and matches still fewer figures differing in overall shape only, then he should make more errors with figures which differ in mouth and eyes, fewer where mouth and shape differ, and virtually none where eyes and shape differ (first row, Table XVI). The extent to which such consistency obtains, for all subject groups, is shown in Table XVII.

Table XVII

Group	Consistent orders	Inconsistent orders	Not relevant	p value
I	10	1	6	.006
II	2	2	5	.673
III	6	0	2	.016
IV	7	3	3	.172
V	8	1	1	.020
Total	33	7	17	.00005

Number of consistent and inconsistent orders for each group. P values from binomial test.

The "not relevant" column gives the number of subjects to whose responses the hypothesis could not be applied, either because they made only/

only one or two errors, if any, or because their errors involved only figures differing in one attribute. This latter group accounts for the discrepancy between the sums of these frequencies and those given in the third column of Table XVI. These subjects show a high level of performance in that they resist the suggestion that any of the non-identical figures are the same as the standard, or else restrict their errors to figures differing minimally, that is by one attribute only. The p values are derived from the application of the binomial test to the distribution of consistent and inconsistent orders, omitting the irrelevant sets of results. Although the individual groups show wide variations in the level of significance associated with the relative frequency of occurrence of the two types of order, there can be no doubt that distribution of type of similarity order for all groups is non-random.

The first of the Tversky-Krantz hypotheses, that dissimilarity is constant over all stimulus pairs differing by one, two or three attributes when the remainder have the same value, cannot be tested fully in the type of situation used here where an absolute judgement of each stimulus pair is required of the subject and observations are not repeated. The other two hypotheses have, however, been shown to hold; as the number of points of difference between two figures increases, the probability that they/

they will not be judged to be the same also increases, and the order of similarity of figure pairs differing in one attribute has been shown at a high level of significance to determine the order of similarity of figures differing in two. It has also been seen (Table XVI) that not all attributes are equally discriminable for all subjects. Although there is a strong tendency for the shape of the mouth to be disregarded as a differentiating attribute by most subjects, this is by no means true for all of them.

Turning now to the analysis of results from set 2 (circle figures), Table XVIII gives the basic data for all groups.

Table XVIII

Group	First response correct		Total response correct		Total response correct as prop. of max. poss.	Total responses	Average spont. MR/ subject
I	119	(78)	139	(54)	91	256	91
II	40	(50)	57	(39)	70	146	70
III	19	(26)	51	(29)	71	176	71
IV	57	(49)	76	(35)	65	220	65
V	17	(19)	46	(24)	51	192	51

Correct responses, etc. for all groups on set 2 (circles) Format as for Table IX.

Again the number of omission errors (third column) brings out most clearly the relationship between age and accuracy of response, with the oldest group/

group showing almost complete success and the youngest only giving 50% correct responses. As with set 1, Group 3 give a large number of responses and consequently compare badly with their peers in terms of proportion of responses correct. The overall level of performance is clearly beyond that of category A, and can therefore be taken as reflecting at best some of the regularity provided by the attributes comprising the stimulus set. This can be seen in more detail in Table XIX, which shows that figures differing even by one level on the attribute of shape are very much less likely to be confused than are figures differing in orientation.

Table XIX

Group	Degree of Difference							
	Orientation		Shape		Both			
	1	2	1	2	S_1O_1	S_1O_2	S_2O_1	S_2O_2
I	52 (53)	15 (15)	17 (30)	3	11 (1)	3	1	0
II	53 (61)	20 (17)	17 (20)	1	8 (2)	1	0	0
III	48 (51)	25 (23)	7 (4)	4	8 (9)	3	3	1
IV	40 (50)	13 (13)	12 (17)	3	22 (13)	6	4	1
V	34 (30)	12 (11)	16 (15)	5	12 (11)	7	8	5

Distribution of wrong responses on set 2. Format as for Table IX.

The lack of success with which orientation differences are detected is still/

still more apparent in Table XX, which corrects for differences in the number of pairs of figures in each of the five classes differentiated by combination of attribute and level. The first row of the table gives the number of stimulus pairs differing to a given degree; the figures in the body of the table represent the number of incorrect responses given to each class of distinct figures as a proportion of the possible total of such responses. The latter figure is arrived at by doubling the number of stimulus pairs in each class and multiplying the result by the number of subjects in each group. The initial doubling of number of stimulus pairs is required by the fact that every member of the stimulus set is matched with every other, so that subjects have the opportunity not only to confuse A with B, but also to confuse B with A.

Table XX

	Degree of Difference							
	Orientation		Shape		Both			
	1	2	1	2	S_1O_1	S_1O_2	S_2O_1	S_2O_2
No. of prs. in each class	6	3	6	3	8	4	4	2
I	30	17	10	3	5	2	1	0
II	44	34	14	4	5	1	0	0
III	60	62	9	5	8	3	3	3
IV	36	25	11	5	13	8	5	4
V	41	30	20	7	11	14	15	13

Number of incorrect judgements of stimulus pairs of each degree of difference as proportion of maximum possible number of such confusions.

The extent to which orientation is ignored in making similarity judgements in comparison with shape is much more clearly brought out here. For all groups, proportionately fewer errors are made with figures differing only by one level of shape than are made with figures differing in two levels of orientation. In both classes of figures differing in only one attribute the hypothesis that judged similarity reflects degree of difference is upheld; in only one case (Group III, orientation) is a greater proportion of errors made with figures differing by two levels than by one, and the discrepancy is small. Errors where both attributes vary are uniformly low; it is only here that differences in type of error among groups becomes clear, in that Groups IV and V give a larger proportion of their responses to these more obviously distinct figure pairs. The group results accurately reflect the data produced by individual subjects; no violations of the ordering hypothesis were produced.

Again an MDS analysis is presented only for Group I. Table XXI shows that the Euclidean metric gives a slightly more accurate configuration, but the differences are not so great that it would be legitimate to exclude the city block metric completely.

Table XXI

	Dimensions		
	2	3	4
Euclidean	.003	.002	.0003
City block	.004	.004	.002

Values of Delta for different metrics,
Group I data.

However, since the values of Delta indicate that the distances in the Euclidean configuration fit the similarity ordering marginally better, it seems more appropriate to decide in favour of it. Nevertheless this finding is in line with the suggestion that the Euclidean metric is more appropriately used where stimuli are composed of integral attributes, as here. In view of the minimal difference between the goodness of fit of the two- and three-dimensional configurations, the former are shown in Fig. IX. That based on the Euclidean metric gives a much closer approximation to the attribute structure of the total set in its regularity, while the other, with the figures clustered in terms of their shape, corresponds more to the suggestion based on the previous analysis that orientation is less likely to be used in discriminating between figures. However, even with the Euclidean metric the pairs of figures differing in orientation are in general more widely separated than those differing in shape. Thus again it is evident that the structure of the stimulus set is being taken account of by the subjects.

The HCS for all groups are given in Fig. X. It is interesting that with the exception of Group V, who have consistently shown less accuracy and systematic use of attribute structure than any other group in the results reported up to this point, none of the hierarchies show the three sets of figures differing in orientation only being brought together before clusters involving figures differing in shape. This may well reflect the small number of errors made by the older groups with figure pairs differing in more than one/

one attribute. As an HCS is built up, the larger entries in the similarity matrix are used first and the higher-level clusterings, which reflect a lower degree of similarity, are frequently based on a relatively small number of confusions. Thus where errors are few and confined to a limited subset of figure pairs, as with the older groups here, the information on which the higher levels of the HCS are based may well be unreliable. This finding does show, in a striking way, the reflection of stimulus structure in responses; although Group V gave only half the possible number of correct responses, their errors are quite evidently not random.

One common feature of the HCS for all groups is that stimulus figures 3, 4 and 8, the three with the smallest size of gap, cluster together at an early stage, and with the exception of Group III, are not linked with any other figures until the final cluster, where the whole set is merged. This indicates that these three figures form a more discriminable subset than any of the others, the two sets of figures with larger gaps being (relatively) more confusing.

It was suggested in the introductory section of this chapter that more asymmetries would be found with set 2 than with set 1, on the grounds that orientation would be utilised in a way distinct from other attributes and that figures in a particular orientation would have matches made to them more frequently than they would be matched to other figures. This was not the case. A few asymmetries occurred in the results for individual groups, but the response totals involved were too small to accept them as reliable;/

reliable; nor was there any consistency across groups in the asymmetries found. Combining data for all groups produced five asymmetries for set 1 and four for set 2, with asymmetry being defined as at least twice as many matches being made of A to B as of B to A. These are shown in Table XXII.

Table XXII

Set 1		Set 2	
1*	6	1*	6
7*	1	7*	1
2*	3	7*	2
2*	8	4*	8
5*	3		

Asymmetries. "*" stands for "is matched more often to."

As can be seen, there are no consistencies here; the hypothesis therefore remains open to further test.

Conclusions:

Accuracy and consistency of response has again been shown to be a function of age, with younger subjects making more omission errors, giving proportionately less correct responses, and making more confusions between figures differing in more than one attribute. The major aspect of the/

the study, however, the use of a normative model to specify the stimulus sets in such a way that the nature of errors can be accurately defined, has been amply justified. The extent to which response patterns have been shown to reflect stimulus structure, using different methods of analysis, indicates that similarity judgements in children are more systematic than has hitherto been realised. The results show that even at four years of age children are capable of detecting and using information about the structure of the total stimulus set, when this is available to them, in a regular fashion. We are now able to suggest an answer to the question raised in the first part of this chapter, on the nature of the decision rule used by subjects in allotting figures to the match/mismatch categories. It was suggested there that if the decision criteria operated in terms of a particular level of similarity, for example any two attributes in common between two figures being enough to define them as identical, then all pairs differing by more than the amount specified would be discriminated, but no others would. The alternative possibility suggested was that certain attributes would be less discriminable than others and that therefore, for a given level of physical similarity, the difference between some pairs of figures would be detected but would not for others. It has been found that certain differences are more likely to be ignored than others, for both stimulus sets used here. The conclusion is therefore that performance is limited by ability to detect features of difference, rather than the application of a deviant notion of what is implied by "the same as."

The/

The fact that certain attributes were consistently ignored here makes the question of the integrality or separability of the stimuli redundant. However they may be conceptualised as entities in themselves, the subjects' redefinition of them is going to nullify any descriptive efforts along these lines. Thus the question of the appropriate metric for the analysis of such data, where the interpretation of the stimulus set from which it is derived can be shown to deviate from the intended structure, as suggested earlier also becomes somewhat of an academic exercise. There was some indication that the Euclidean metric was more appropriate for the set 2 data, as would have been expected; but why this should be so when one attribute was so consistently overlooked is not clear.

Experiment IV: Form discrimination using pair comparison presentation

This study was intended to show what differences, if any, a different form of presentation of the figures to be discriminated would make to the findings. The material of set 1 (schematic faces) was used in a pair comparison task. All pairs of figures were presented to the subjects, one at a time; the instructions simply took the form of a question; "Is this (point to one stimulus) the same as this?" (point to other). As far as was possible the position of each figure, to the left or right, was equated over its appearances. Order of presentation of pairs was random. There were 28 distinct pairs and 8 identical ones. Twenty subjects were used in all, ten from Group IV and the ten of Group V. The task was administered to all subjects some weeks after they had completed Experiment III.

No subject failed to describe a physically identical pair of figures as being the same. Immediately there is a considerable difference from the findings of Experiment III, where both of these groups made a number of omission errors. Two possible reasons could be advanced for the discrepancy. By presenting figures two at a time the child's attention is being focussed sharply on the relevant aspects of the situation; the distraction provided by a large comparison array, with the possibility that poor scanning will result in his literally not seeing the correct match, is eliminated. Also the form of question used, where the child has the difficult/

difficult task of saying "No" if he is to be correct on most trials, may have produced a response bias which had little to do with the stimulus as such.

Although both of these contaminating factors may have been operating, they cannot account for the distribution of errors. This is shown in Table XXII.

Table XXIII

Group	Degree of Difference												Total
	1 attribute				2 attributes				3 attributes				
	S	E	M		SE	SM	EM		SEM				
IV	11	17	14	(39)	18	15	20	(49)	14	(12)			109
V	15	26	24	(39)	27	25	23	(47)	20	(14)			160

Distribution of errors in pair comparison task. Bracketed figures, number of errors in each class as proportion of total number of errors.

The consistency found in the previous experiment has vanished; a greater proportion of errors is made with stimulus pairs differing in two attributes than with those differing in one, and a far greater proportion of three-attribute differences between figures result in errors. The order of errors for figures differing in two attributes follows that for one-attribute differences for Group IV, just, but not for Group V or the two combined. Thus/

Thus it appears that these subjects have answered "No" almost at random; certainly the pattern of response is very much less systematic than would have been anticipated on the basis of the findings from the last experiment.

This is shown again in Fig. XI, in the HCS for the two groups. The initial clusters differ in either two or three attributes; the regular partitioning of the stimulus set along the lines given by its structure, observed in Experiment 2, has been quite lost, as can be seen by a comparison of Fig. XI and Fig. XVIII.

The contrast between these findings and those of Experiment III indicates beyond doubt that only when the whole stimulus set is available to the child will his responses reflect its structure. Had this experiment been carried out in isolation it would have been legitimate to suggest not only this was not possible for children of this age, but that there was some doubt as to whether they were capable of comprehending and using the notion of "same as." Context effects are well known in psychology; but the magnitude of the effect that the context of the total set can have in a form discrimination task of this nature is surprising, to say the least.

Chapter 5

Concluding comments

The studies reported here have been exploratory in nature, designed to answer questions of the "what-happens-if?" variety. Putting the results of the longitudinal and cross-sectional experiments together, the picture of the development of form discrimination skills which emerges is close to that described by E.J. Gibson. Improvement in performance does reflect an increasing awareness of stimulus features and a more refined use of them. The developmental process is not straightforward, however. Starting from a level where responses reflect situational rather than stimulus factors, which is to a large extent defined by the complexity of the problems presented, the child moves on to a stage where it is necessary for him to match a number of comparison figures to the standard because he is not yet capable of detecting all the features which define the single correct response. This is the final level of competence. We have seen that it is failure to pick up relevant information that is responsible for the intermediate stage, rather than the use of a criterion of partial identity between figures in coming to the match/mismatch decision, from the fact that stimulus figures differing by the same number of attributes are not confused to the same extent. In addition, the homomorphism shown between/

between the structure available in the stimulus set and that of responses indicates a systematic misuse of information which cannot be accounted for simply in terms of detection of partial identity as such.

These comments only apply to results derived from a matching-from-sample task. When pair comparison presentation is used the stimulus structure is no longer detected, and although the probability that fully identical figure pairs will be correctly responded to increases markedly, errors become random. It is quite likely that the use of a discrimination learning paradigm would bring about an awareness of the stimulus structure; but the point here is simply that with a matching task the whole stimulus set has to be made available if we wish to study the child's ability to detect regularity.

Perhaps the major implication of the findings reported here is that the nature of the stimulus has to be considered with some care before the use to which the subject puts it can be adequately discussed. This is less important if all that is required is information on how well different figures are differentiated under particular circumstances; but the field of investigation must be expanded and made more rigorous, by the use of an appropriate normative model, if we are to be able to go beyond this and discuss strategies and processes underlying performance. Here a complete specification of the structure of the stimulus set is essential; without the baseline provided by such a description, it is not possible to analyse/

analyse errors systematically and see how they relate to the defined differences among the members of the set, especially since children are so ready to redefine the stimuli presented to them. But when these factors are taken into consideration, much more can be said about how the child goes about detecting regularity, order and structure, how well he can do this in the context of form discrimination experiments, and what kinds of skills he brings to the task.

BIBLIOGRAPHY

Attneave, F. and Arnoult, M.D. (1956) The quantitative study of shape and pattern perception. Psychol. Bull. 53, 452 - 471.

Beals, R., Krantz, D.H. and Tversky, A. (1968) Foundations of multidimensional scaling. Psychol. Rev. 75, 127 - 142.

Birch, H.G. and Lefford, A. (1967) Visual Differentiation, Intersensory.

Integration and Voluntary Motor Control. Lafayette, Society for Research in Child Development, Mon. vol. 32, No. 2.

Braine, L.G. (1965) Age change in mode of perceiving geometric forms. Psychonom. Science, 2, 155 - 156.

Behrman, B.W. and Brown, D.R. (1968) Multidimensional scaling of form; a psychophysical analysis. Perception and Psychophysics, 4, 19 - 25.

Brown, D.R. and Andrews, M.H. (1968) Visual form discrimination: multidimensional analyses. Perception and Psychophysics, 3, 401 - 406.

Brown, D.R. and Goldstein, J.A. (1967) Stimulus correlates of the discrimination behaviour of children. Psychonom. Science, 9, 177 - 178.

Brown, D.R. and Owen, D.H. (1967) The metrics of visual form; Methodological Dyspepsia. Psychol. Bull. 68, 243 - 259.

Bryant, P. E. (1969) Perception and memory of the orientation of visually presented stimuli by young children. Nature, Lond. 224, 1132 - 1133.

Campbell, R.N. and Young, B.M. (1968) Some aspects of classifactory behaviour in pre-school children. (Presented to the Education Section of the British Psychological Society).

Donaldson, M.C. (1963) A Study of Children's Thinking. London, Tavistock.

Donaldson, M.C. and Wales, R.J. (1968) On the acquisition of some relational terms. In Hayes, J.R. (Ed.) Proceedings of the 1968 Carnegie Symposium on Cognition. New York, Wiley.

Doran/

Doran, J.E. and Hodson, F.R. (1966) A digital computer analysis of palaolithic flint assemblages. Nature, Lond. 210, 688 - 689.

Galanter, E.H. (1956) An axiomatic and experimental study of sensory order and measure. Psychol. Rev. 63, 16 - 28.

Garner, W.R. (1966) To perceive is to know. Amer. Psychologist, 21, 11 - 19.

Garner, W.R. (1970) The stimulus in information processing. Amer. Psychologist, 25, 350 - 358.

Garner, W.R. and Morton, J. (1969) Perceptual independence: Definitions, models and experimental paradigms. Psychol. Bull. 72, 233 - 259.

Ghent, L. (1961) Form and its orientation: a child's-eye view. Amer. J. Psychol. 74, 177 - 190.

Ghent, L. and Bernstein, L. (1961) Influence of orientation of geometric forms on their recognition by children. Percept. Mot. Skills, 12, 95 - 101.

Gibson, E.J. (1965) Learning to read. Science, 148, 1066 - 1072.

Gibson, E.J. (1969) Principles of Perceptual Learning and Development. New York, Appleton-Century-Crofts.

Gibson, E.J., Gibson, J.J., Pick, A. and Osser, H. (1962) A developmental study of the discrimination of letter like forms. J. comp. physiol. Psychol. 55, 897 - 906.

Gibson, E.J., Osser, H., Schiff, W. and Smith, J. (1964). A basic research program on reading. Final Report on Cooperative Research Project No. 639 to the Office of Education, Department of Health, Education and Welfare.

Gibson, J.J. (1966) The Senses Considered as Perceptual Systems. Boston, Houghton Mifflin.

Griffiths, J.A., Shantz, C.A. and Sigel, I.A. (1967) A methodological problem in conservation studies: the use of relational terms. Child Development, 38, 841 - 848.

Huttenlocher, J. (1967) Discrimination of figure orientation: effects of relative position. J. comp. physiol. Psychol., 63, 359 - 361.

Handel, S. and Garner, W.R. (1966) The structure of visual pattern associates and pattern goodness. Perception and Psychophysics, 1, 33 - 38.

Imai, S. and Garner, W.R. (1965) Discriminability and preference for attributes in free and constrained classification. J. exper. Psychol., 69, 596 - 608.

Imai, S. and Garner, W.R. (1968) Structure in perceptual classification. Psychonomic Monograph Supplements Vol. 2, No. 9 (whole No. 25).

Jeffrey, W. E. (1966) Discrimination of oblique lines by children. J. comp. physiol. Psychol., 62, 154 - 156.

Johnson, S. C. (1967) Hierarchical clustering schemes. Psychometrika, 32, 241 - 254.

Kerpelman, L.C. and Pollack, R. H. (1964) Developmental changes in the location of form discrimination cues. Percept. Mot. Skills, 19, 375 - 382.

Kruskal, J. B. (1964 a,b) Multidimensional scaling by optimising goodness of fit to a non-metric hypothesis. Psychometrika, 29, 1 - 27 and 115 - 129.

Michels, K.M. and Zusne, L. (1965) Metrics of visual form. Psychol. Bull. 63, 74 - 86.

Over, R. and Over, J. (1967) Detection and recognition of mirror-image obliques by young children. J. comp. physiol. Psychol. 64, 467 - 470.

Pick, A. D. (1965) Improvement of visual and tactual form discrimination. J. exp. Psychol., 69, 331 - 339.

Rosenblith, J. F. (1965) Judgements of simple geometric forms by children. Percept. Mot. Skills, 21, 947 - 990.

Royer, F.L. and Garner, W.R. (1966) Response uncertainty and perceptual difficulty of auditory temporal patterns. Perception and Psychophysics, 1, 41 - 47

Rudel/

Rudel, R.G. and Teuber, H.L. (1963) Discrimination of direction of line in children. J. comp. physiol Psychol., 56, 892 - 898.

Schwartz, J. (1962) The pernicious influence of mathematics on science. In, Logic, Methodology and the Philosophy of Science, ed. Nagal, E. Suppes, P. and Tarski, A. Stanford.

Sekuler, R.W. and Houlihan, K. (1968) Discrimination of mirror images: choice time analysis of human adult performance. Quart. J. exp. Psychol. 20, 204 - 207.

Sekuler, R.W. and Rosenblith, J.F. (1964) Discrimination of direction of line and the effect of stimulus alignment. Psychonom. Sci., 1, 143 - 144.

Shepard, R.N. (1962 a,b) The analysis of proximities: multidimensional scaling with an unknown distance function. Psychometrika, 27, 125 - 140 and 219 - 246.

Shepard, R.N. (1964) Attention and the metric structure of the stimulus space. J. math. Psychol., 1, 54 - 87.

Stevens, S. S. (1966) On the operation known as judgement. Amer. Scientist, 54. 385 - 401.

Taylor, J.A. and Wales, R.J. (1970) A developmental studyoof form discrimination in pre-school children. Q.J. exp. Psychol. 22, 720 - 734.

Torgerson, W.S. (1965) Multidimensional scaling of similarity. Psychometrika, 30, 379 - 393.

Torgerson, W.S. (1958) Theory and Methods of Scaling. New York, Wiley.

Tversky, A. and Krantz, D. (1969a) The dimensional representation and the metric structure of similarity data. Michigan Mathematical Psychology Program, 69 - 7.

Tversky, A. and Krantz, D. (1969b) Similarity of schematic faces: a test of interdimensional additivity. Perception and Psychophysics, 5, 124 - 128.

Vurpillot, E. (1968) The development of scanning strategies and their relation to visual differentiation. J. exp. child Psychol., 6, 622 - 650.

Wohlwill/

Wohlwill, J.F. and Wiener, M. (1964) Discrimination of form orientation in young children. Child Dev. 35, 1113 - 1125.

Zaporozhets, A. V. (1964) The development of perception in the preschool child. In, European Research in Child Development, ed. Mussen, P. H. Society for Research in Child Development, Mon. Vol. 30, serial No. 100.

Q. Jl exp. Psychol. (1970) **22**, 720–734.

A DEVELOPMENTAL STUDY OF FORM DISCRIMINATION IN PRE-SCHOOL CHILDREN

JOHN A. TAYLOR AND ROGER J. WALES
Department of Psychology, University of Edinburgh

Using a matching-from-sample technique the discriminative ability of a group of pre-school children was tested twice, at mean ages 3–8 and 4–9, in relation to an independent test of their comprehension of the notion "same". A sequence of three stages is described, for both shape and orientation discrimination; in the first, characterized by a large number of errors, the child appears to respond largely in terms of proximity of a comparison figure to the standard; in the second his responses reflect gross aspects of similarity between the figures in the comparison set, in that several matches are made to one standard all of which have certain attributes in common with it. Finally correct and unique choices are made. Attributes used in differentiating figures are described, and a number of theoretical and methodological problems are discussed.

Introduction

Previous work on form discrimination shows that the ability to differentiate between different forms is well developed in normal children by the age of 5, whether the stimuli are simple geometric figures (Birch and Lefford, 1967, chapter 4), random polygons of the Attneave Arnoult type (Brown and Goldstein, 1967), or letters of the alphabet (Gibson, Osser, Schiff and Smith, 1964). The discrimination of identical forms in different orientations is a less easy task. Ghent has demonstrated definite orientation preferences, affecting accuracy of recognition in a matching from sample situation, in children up to five years old (Ghent, 1960; Ghent and Bernstein, 1961; Braine, 1965). Stimulus alignment has been shown to influence responses in a regular fashion in children (Sekuler and Rosenblith, 1964; Huttenlocher, 1967*a*, *b*), and in adults (Sekuler and Houlihan, 1968). Making due allowance for this complicating factor (which has been demonstrated only with pair comparison presentation and may therefore not be generalizable to larger arrays) it has been generally found that left-right reversals (e.g. b–d) are more difficult than up-down reversals (e.g. b–p), with 90° rotation and 180° rotation (e.g. b–q) following in that order (Huttenlocher, 1967*a*; Rosenblith, 1965; Rubel and Teuber, 1963; Wohlwill and Weiner, 1964). The last-mentioned authors also point out that latencies are longer for low-directionality than for high-directionality figures, which can be taken as indicating a greater degree of "effort", in some sense, being required when orientation is not clearly marked. Gibson *et al.* (1964) report a similar increase in latencies in the matching of highly confusable letters.

The well-known series of papers by Eleanor Gibson and her colleagues (Gibson, Gibson Pick and Osser, 1962; Gibson *et al.*,1964; Gibson, 1965) presents a more

Reprinted from The Quarterly Journal of Experimental Psychology
Vol. XXII, Part 4, November, 1970

detailed and larger-scale analysis of a particular aspect of this problem as it appears in the first stages of learning to read: "... what dimensions of difference must a child learn to discriminate for each letter to become unique?" (Gibson *et al.*, 1964, p. 2). In the first paper cited they found that certain transformations of letter-like forms were more likely to be confused with the standard form, and that the probability of obtaining correct responses with this type of material increased with age. Specifically, transformations which altered the shape of the comparison figure in relation to the standard, either by making a gap in a closed figure or by the closure of an open one, produced relatively little confusion even in 4-year-old subjects. There were two types of transformation which preserved the original shape of the figures; perspective, which essentially produced slight size changes; and rotation and reversal, which affected orientation. There was little improvement in ability to discriminate the former in subjects up to the age of 8, while errors on the latter fell steadily to virtually zero by the same age. There are certain apparent discrepancies between the overall findings here and the general ones mentioned above, which Rosenblith (1965) has resolved by a re-analysis of the relevant parts of the data.

However, none of these studies takes the child's conception of similarity into account. This is of some importance, since unless his interpretation of the usual form of instruction "Find a figure in this array which is 'the same as' the standard" is known, it is not possible to decide if his errors are due to a genuine failure of discrimination or to his not possessing adult criteria of similarity. Garner (1966) has pointed out that discrimination is a process akin to classification, in which responses are determined more by the properties of sets of stimuli rather than by those of individual members of the set. This view implies that we must know to what extent the child's understanding both of the relations expressed by "the same as", which is the basis of the judgement underlying his discriminative response, and of the aspects of the stimulus situation to which he applies it, correspond to those of adults. By allowing our subjects to define the subsets of stimulus figures which they consider to be equivalent we can see directly what attributes they take to be relevant in assessing the stimuli, and also infer how widely and how appropriately they are applying their notions of similarity.

Our concern here is with the interaction of two aspects of functioning; the nature of the rules or principles of organization the child has available to him for classifying, and so imposing some order or structure on, his world, and the way in which these are applied to the information that he extracts, in a more or less adequate manner, from the material presented to him.

There are two possibilities for error in the child's understanding and use of "same"; either the reference to identity of some kind, the "meaning of the word", is not known, or, while it is understood to refer to things which are in some sense identical the notion of identity does not correspond to adult criteria. From the results of a classification test administered to our group of subjects at approximately the same times as the discrimination task it appears that, with two exceptions which will be discussed later, they could respond appropriately when asked to find an object "the same in some way" as one picked out by the experimenter. This test is described more fully in Campbell and Young (1968), and its linguistic aspects discussed by Donaldson and Wales (1968). Thus we can say that errors are nc due to any gross failure of understanding. Our subjects were tested twice; thes comments apply to them at the time of the first testing, since they were all able t perform this (classification) task equally well by the time of the second testing.

It should be pointed out that this description of the child's ability to use the notio of "same" applies only to the situation where he is asked to find an object o picture in an array which is "just the same" as some standard object or picture whic is separate from the array. The results of another test administered to this sam group of children at approximately the same time as the classification and discrimin ation tasks indicate that they do not make correct responses when presented wit two pictures pictures side by side and asked if one is the same as the other. Eac of the pictures was of a string of four beads, presented vertically; one bead on on string was of a different colour from the corresponding bead on the other string Stimulus complexity may therefore be an important factor in this apparent discrep ancy.

In sum, this investigation is concerned with the kinds of stimulus propertie used by children to discriminate between figures. In addition, we wish to fin out how the at least rudimentary notion of what is implied by "same" possessed b our subjects is applied to this type of problem, and how, if at all, its applicatio changes in terms of the subset of equivalent figures defined by the child, and of th attributes of stimulation used as a basis for differentiation at different ages.

Method

The three series of stimulus figures shown in Fig. 1 were chosen. Series I, coverin orientation, represented a set of problems of increasing difficulty as measured by the numbe of errors made during pilot testing. This was intended as the main index of change betwee first and second administrations, as we found shape discrimination to be very much simple for younger subjects. Series II and III were both intended to provide information o shape discrimination in relation to both shape and orientation differences, but from distinc points of view. (Since children in our subjects' age group never, during preliminar testing, confused rectilinear and curvilinear figures, we did not test this discriminatio here.) In Series II we attempted to provide as many stimulus attributes as possible, whi retaining an array of manageable size, to see what kinds of confusions were made whe figures varied in shape, orientation, number of sides, and in being open or closed. Serie III represents a more systematic approach to the same question; apart from the two close figures, which were retained as a further check on Series II findings differ only in orien ation and degree of openness. The object here was to see if, given a certain regularity i stimulus attributes, this would be reflected in responses in any systematic fashion.

For three reasons we decided to use a matching-from sample technique, using fairly larg arrays, rather than pair comparison, three-choice oddity problems, or matching from sma samples. The first, the definition by the subjects of those subsets of the total stimulus s which they considered to be equivalent, has already been discussed. This is relatec second, to the way in which "same" is understood; it is much easier to test the limits of th notion by direct observation of the groups of figures the child describes as being the sam than to infer them from single responses to stimulus pairs or triplets. Finally, there is th mechanical problem of number of stimulus presentations; we wanted to study form an orientation discrimination separately and also to see what happened when both form an orientation differences were combined in the one array, and so had to use an economic method of gathering data. A further consideration arose during pilot testing of our materia part of which involved pair comparison presentation. We found here, particularly wit

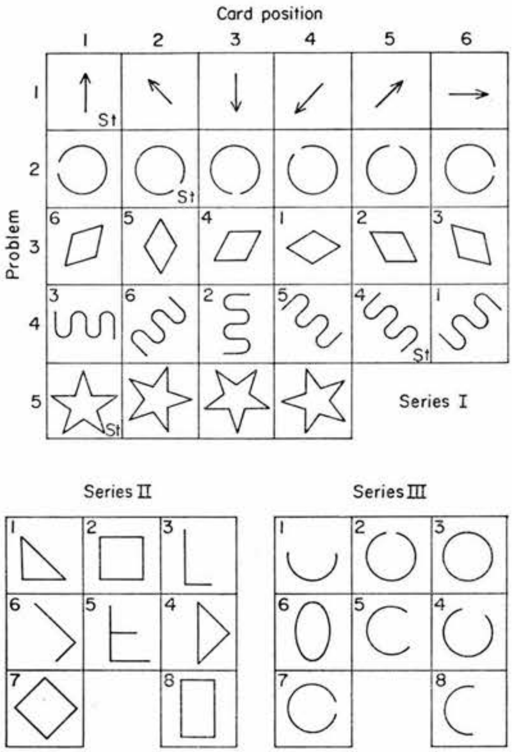

FIGURE 1. Experimental material. In Series I standards are shown by St; numbers shown with figures in problems 3 and 4 give the order of layout of the comparison array on second presentation. Numbers in Series II and III are figure labels.

figures differing only in orientation, that the child tended to label the figure and make his judgement in terms of the label (the question asked was, "Is this picture just the same as this one?" accompanied by appropriate pointing by E). For example, when faced with two ellipses in different orientations a typical response to this question would be, "That's an egg and that's an egg" or "Both eggs", rather than a direct "yes" or "no". Although this labelling also occured with the larger arrays finally used in the experiment proper, particularly with the younger children, it was much easier to get the child to accept that there was one that was the same as the standard.

Subjects

These comprised the 24 children participating in an intensive longitudinal study of cognitive and linguistic development in pre-school children. They were tested twice, at mean age 3 years 8 months (range 3–4 to 4–0) and 4 years 9 months. The earlier sample included one child who was not available for later testing. Thus there were 13 male and 11 female subjects at the time of first testing, and 23, the same group less one boy, at the second. The two testing sessions will hereafter be referred to as pretest and post-test, respectively.

Presentation

The three series of figures were drawn in black ink on 3-in^2 white cards, which were covered with washable plastic. During each testing session the experimenter and subject sat at a small table, appropriate to the size of the Subject, with the experimenter on the subjects' right. Series I problems were presented one at a time, the experimenter laying out the comparison array as shown in Fig. 1 with a duplicate of the standard just below the centre of the array. The experimenter pointed to the standard, told the subject to have a good look at it, then said, "See if you can find a picture up here that's just the same as this one", running his finger slowly the length of the line of comparison figures while doing so, and pointing again at the standard at the words ". . . this one". Three different problems were used to introduce the task; with these, errors were explained and correct matches indicated. This was not done with any of the experimental material; the experimenter made general approving comments, such as ,"Good, that's the way", whatever the response, irregularly throughout each session. The arrays for Series II and III were laid out as shown in Fig. 1, duplicate copies of the standard being placed in the gap in the bottom row. Instructions were as for Series I, with no introductory examples and no correction of errors. Order of presentation of the standards was the same for all subjects, but bore no relation to the order of figures in the array. The three series were presented in the order of their numbering, problems 3 and 4 of Series I being presented again after Series III, with the comparison cards laid out in the alternative order shown.

Questions of attention, interest, and general willingness to perform are of central concern in any investigation with children of this age. All our subjects were quite happy to "come and play games" or "do some puzzles". Time for a full session varied from 15 to 20 min., depending on the amount of time taken up by irrelevancies between problems, but care was taken to ensure that subjects attended to each problem as it was presented.

Results

We shall first present an analysis of the results for each series individually, and then give an overview of our subjects' performance over the three parts of the experiment.

Series I

The results for this series are shown in Tables I and II. Two subjects in the pre-test gave card 1 as the match to the standard in all problems; their results are omitted from Table I. Where the number of first choices shown is less than the number of subjects, this indicates that one or more of the subjects matched all

TABLE I

Series I pre-test matches

Card Position	Problem 1		2		3		3R		4		4R		5	
1	*15*	*13*	6	5	5	2	7	5	1	0	6	3	*10*	*8*
2	1	1	*7*	*4*	4	3	*4*	*1*	6	5	1	0	2	1
3	6	6	4	2	10	7	5	5	3	2	2	2	8	7
4	2	1	10	7	6	2	8	8	13	11	*15*	*13*	6	5
5	1	0	3	0	*6*	*3*	2	1	*8*	*3*	5	3	–	–
6	2	1	7	3	8	5	2	1	5	1	4	1	–	–
Σ responses	27	22	37	21	39	22	28	21	36	22	33	22	26	22
D_{max}	0·39†	0·42†	0·06	0·19	0·10	0·11	0·19	0·24	0·22	0·18	0·23	0·37	0·14	0·13

Number of first responses to each problem is given in right-hand columns; total number of responses in left-hand columns. Standards are italicized. D_{max} values are from Kolmogorov-Smirnov one-sample test.

N=22; total correct first responses=45.

† P 0·01; Σ total responses = 226.

TABLE II

Series I post-test matches

Card Position	Problem 1		2		3		3R		4		4R		5	
1	*22*	*22*	3	1	7	2	4	2	2	0	5	1	*14*	*12*
2	1	1	*14*	*11*	2	1	*11*	*5*	10	4	2	0	7	3
3	1	0	7	5	9	4	15	8	3	1	1	0	9	4
4	0	0	6	4	7	5	9	6	14	12	*20*	*19*	4	1
5	0	0	2	0	*14*	*6*	2	0	*15*	*5*	10	2	–	–
6	1	0	3	2	2	4	8	1	8	0	4	0	–	–
Σ responses	25	23	35	23	51	22	49	22	52	22	42	22	34	20
D_{max}	0·71†	0·79†	0·15	0·25	0·18	0·20	0·13	0·29‡	0·21	0·27	0·31†	0·46†	0·16	0·35‡

Number of first responses to each problem is given in right-hand columns, total number of responses in left-hand columns. Standards are italicized. D_{max} values are from Kolmogorov-Smirnov one-sample test.

$N=22$; total correct first responses $=45$.

† $P < 0·01$; ‡ $P < 0·05$; Σ total responses $=286$.

the comparison cards, unshakably, to the standard, without singling out any one as a preferred choice even after questioning.

Improvement over time is shown by the increased number of correct first choices in the post test as compared with the pre-test; 8 subjects failed to repeat their previous success on the same post-test items, but none of these got more than one of the problems correctly answered in the pre-test wrong in the post-test. A Wilcoxon matched-pairs signed-rank test shows the difference to be highly significant ($N = 20$; $T = 32·5$,; $P < 0·005$ (one-tailed)). The Kolmogorov-Smirnov D_{max} values shown also demonstrate this improvement, in terms of an altered distribution of responses; three of the 14 distributions are significantly different from chance in the pre-test, while 6 are in the post-test Greater success occurs with problems 1 and 2, where orientation of the figures is more clearly marked than in any other problem. The large number of correct responses in the repeat of problem 4 is unexpected, but the artificial nature of this result is clear from Table III, which shows the presence of a strong positioned bias. This is done by comparing the number of wrong choices of a figure in a particular position in the array, as a match to the standard, with the total number of times a choice of a figure in that position would have been correct. The numbers in Table III refer to the number of choices of a figure in a given position, totalled over all problems except 5. The order of preference is the same, although in different degrees, for both pre- and post-test; a figure in position 4 is most likely to be chosen, followed by one, in position 3, 1, and 6, in that order.

Series II

A multidimensional scaling analysis was carried out on the results from Series II and III. The object of this technique is to find a spatial configuration, in as few dimensions as possible, in which the judged similarity of the members of a set of stimulus objects is reflected in their proximity in the plot. In this case, the more

TABLE III

	Card position	1	2	3	4	5	6
Pre-test	Possible correct choices	22	42	0	22	44	0
	Total choices	28	14	24	42	10	12
	Wrong choices	15	9	24	29	4	12
Post-test	Possible correct choices	23	45	0	22	44	0
	Total choices	30	22	18	46	15	7
	Wrong choices	8	6	18	27	4	7

Number of first choice matches in Series I given to figures in each position in relation to available correct matches totalled over all problems except 5. Discrepancies between possible correct choices reflect omission of two subjects from the pre-test, and of those subjects giving "all" responses.

similar two figures are seen to be, that is the more often they are confused, the closer together they will be in the final configuration. The method is fully described by Shepherd (1962*a*, *b*) and by Kruskal (1964*a*, *b*); the particular version used here was developed by J. Doran (Doran and Hobson, 1966).

Performance in Series II was good in the pretest and improved considerably in the post-test, as shown by the increased number of correct responses, the drop in total of responses, and the small number of confusions (Table IV). The difference between the number of correct responses is again highly significant by the Wilcoxon test ($N = 16$; $T = 2\cdot5$; $P < 0\cdot005$ (one-tailed)). The scaling configurations, in two dimensions for both pre- and post-test, are shown in Figure 2. The three pairs of figures in different orientations are very closely related to each other. There is a clear separation between open and closed figures, and, within the former group, between three- and four-sided ones. In the post-test these distinctions are more definite than they may appear from the plot, in view of the minimal number of confusions made between figures in these groups. The two "unique" figures, 5 and 8, have the highest percentage of correct responses in both sessions; the corresponding results for the other figures reflect the extent to which the two exemplars of each in different orientations are taken as similar. Figure 8 was matched to all other figures in the pre-test a large number of times; this may well reflect a position

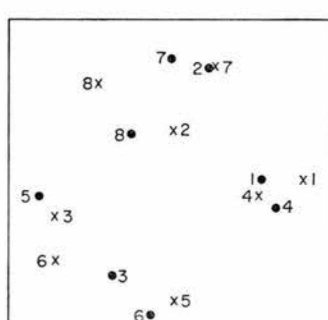

FIGURE 2. Pre- (●) and post-test (×) configurations, Series II. Delta = 0·004 and 0·0006 respectively. The axes merely provide a frame of reference and consequently are unlabelled.

preference effect similar to that already noted in Series I. It is difficult to account for its popularity in other terms.

The kinds of features used in making these judgements may be seen in Table IV, which also gives total number of matches and correct matches to each figure. In the post-test only figure 8 is matched to, or has matched to it, a non-identical figure differing in orientation. Several subjects made comments such as, "That's a wee bit the same" or "That's the same but bigger" when comparing 2 and 8, thus demonstrating that they could differentiate the two while being uncertain whether they should call them "the same". In the pre-test, if we take the choice of 8 as being due to some positional artifact, the only confusions between open and closed figures are the choice of 3 as a match to 4, which was only given by subjects who had already given either 2 or 1 as a match to 4, which was only given by

TABLE IV

Principal matches, total matches and correct matches on Series II

Standard	Pre-test					Post-test			
				Total matches	Correct matches			Total matches	Correct matches
1	1 (40)	4 (31)	2 (11)	35	14	1 (61)	4 (39)	31	19
2	2 (48)	8 (19)	7 (16)	31	15	2 (69)	7 (22)	32	22
3	3 (46)	6 (17)	8 (11)	35	16	3 (76)	6 (17)	29	22
4	4 (48)	1 (25)	3 (8)	40	19	4 (75)	1 (21)	28	21
5	5 (58)	8 (13)	3 (10)	31	18	5 (92)	–	25	23
6	6 (55)	3 (14)	1 (10)	29	16	6 (69)	3 (31)	29	20
7	7 (46)	2 (24)	8 (11)	37	17	7 (82)	2 (18)	27	22
8	8 (70)	1 (10)	2 (7)	30	21	8 (88)	2 (8)	25	22
			Total	268	136		Total	226	171

Bracketed numbers are percentages of total matches made to a given standard; numbers before brackets refer to figures.

subjects who had already given either 2 or 1 as a match to 4, and or 1, 4 and 2 as a match to 6. There seems to be some whole-part effect here, and also in the matching of 1 to 2 and 8, and of 7 to 4, but too much should not be made of this since any one of our sets of figures can be described in terms of parts common to the whole set with the possible exception of 5. Orientation as a distinguishing criterion is to a large extent ignored in both pre- and post-tests.

One surprising feature of these results was the asymmetric nature of both matrices. The major asymmetries are shown in Table V, omitting those in which figure 8 occurs. Two factors must be taken into account here; the degree to which the standard figure is seen as unique, and the number of other figures which are seen as similar to it. Both of these will determine how many "spare" judgements, so to speak, will be available for this kind of effect. Nevertheless, this does show the influence of context in a rather striking way; degree of judged similarity will reflect not some absolute relation between two figures, but the way in which that relation is perceived against the background of the total stimulus set.

TABLE V

Series II asymmetries

Pre-test	Post-test
1 (31) — (25) 4	1 (38) — (21) 4
2 (16) — (24) 7	2 (21) — (18) 7
3 (13) — (17) 6	3 (17) — (31) 6
1 (0) — (10) 6	

A(X) — (Y)B: A is matched to B, Y times; B is matched to A, X times. X and Y numbers are percentages.

Finally, let us consider the two subjects mentioned earlier, whose use of "same" was different from that of the rest of the group. For one, "same in some way" meant identical in all respects, from his classification performance; he made no errors, of either confusion or omission, on this series. For the other it meant non-identical; she got every item wrong, apparently following a rule of her own, if you can't find a figure which is part of the standard then find one of which the standard is a part. Her matches were (standards first); 1 = 2; 2 = 1; 3 = 5; 4 = 2,3,7; 5 = 3; 6 = 4; 7 = 4 ; 8 = 1, 2. Both, however, failed to follow their own rules in the other two pre-test series.

Series III

Overall performance here is not so good as in Series II, in that more confusions are made in both pre- and post-tests. Improvement over time is not so obvious, but is nevertheless demonstrated on the same criteria as used for the previous series. The Wilcoxon test gives $N = 20$, $T = 14$, $P < 0{\cdot}005$ (one-tailed). Three-dimensional configurations were required for both sets of results; these are shown in Figure 3. The two closed figures, 3 and 6, are well separated from the rest in both plots, again demonstrating the lack of confusability between open and closed

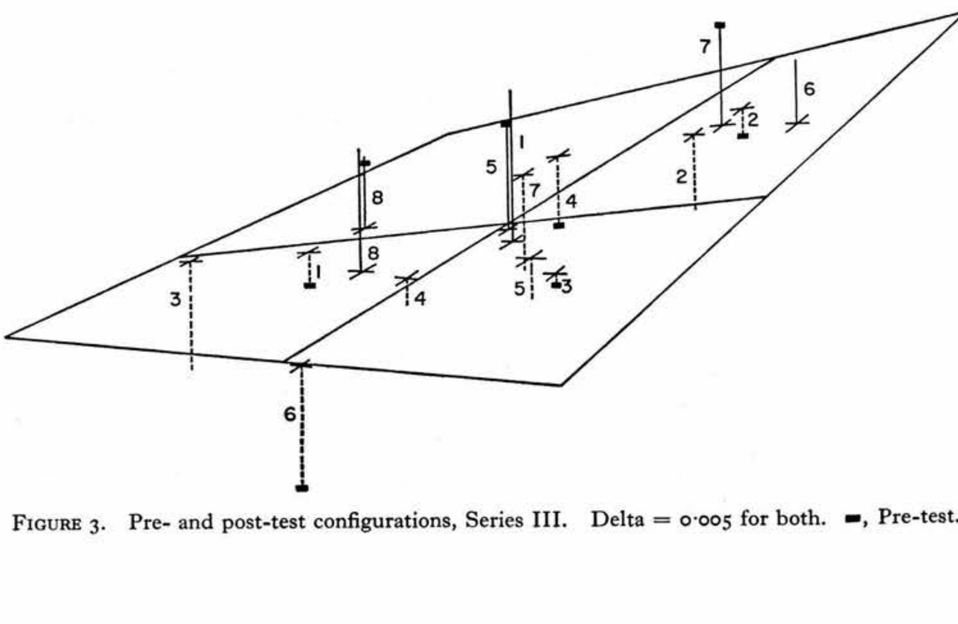

FIGURE 3. Pre- and post-test configurations, Series III. Delta = 0·005 for both. ▬, Pre-test.

TABLE VI

Principal matches, total matches and correct matches on Series III (format as for Table IV)

Standard	Pre-test				Total matches	Correct matches	Post-test			Total matches	Correct matches
1	1 (29)	8 (29)	4 (17)	5 (17)	42	12	1 (46)	8 (27)	4 (11)	37	17
2	2 (38)	7 (25)	4 (13)	5 (10)	32	12	2 (76)	7 (10)	4 (7)	29	22
3	3 (51)	5 (11)	1 (8)	2 (8)	37	19	3(79)	7 (10)	–	29	23
4	4 (28)	8 (19)	2 (16)	1 (13)	32	9	4 (63)	5 (20)	1 (7)	30	19
5	5 (34)	8 (28)	3 (10)	4 (10)	29	10	5 (49)	8 (23)	4 (20)	35	17
6	6 (71)	3 (11)	8 (7)	–	28	20	6 (96)	–	–	24	23
7	7 (50)	2 (19)	–	–	32	16	7 (49)	2 (22)	5 (11)	37	18
8	8 (74)	1 (15)	2 (7)	–	27	20	8 (79)	1 (11)	5 (11)	28	22
				Total	259	118			Total	249	161

figures. In the pre-test, of the remaining six figures, 1, 2, and 4, with gaps uppermost, project in the opposite direction to figures 5, 7 and 8, with gaps at the side. Degree of openness is reflected in the relative proximity of 2 and 7, at least open, and 1 and 8, most open; 4 and 5 are much more widely separated, and seem to indicate an area of uncertainty over the appropriate criteria to be used in matching. These are the most difficult items, with fewer correct responses than any other.

The criterion for matching seems to switch from a roughly equivalent emphasis on shape and orientation in the pre-test to shape only in the post-test. Figure 3 shows 3 and 6 again to be rarely confused with the open figures. 2 and 7, 5 and 4, and 1 and 8, are clustered in pairs projecting in the same direction. Thus orientation is largely ignored as a determinant of similarity, with matches being made on the basis of degree of openness. Table VI shows this very clearly. Responses are not so widely distributed in the post-test, and the order of choice is, with the exception of 5, systematic in a way which is not the case for the pre-test, reflecting the greater likelihood of figures of the same "shape" but different orientation being judged to be the same as the standard.

Asymmetries here are more marked than in Series II (Table VII). With the exception of those between figures 5 and 8, in both sessions, they all obtain between figures opposed in orientation. It is not immediately apparent how this finding should be interpreted. It may be related to the way in which the two attributes of shape and orientation are used in judging the figures. From Table VIII we can see that an equivalent mean percentage of mismatches are made between figures which differ only in their orientation in both pre- and post-tests, but that the number

TABLE VII

Series III asymmetries (format as for Table V)

Pre-test	Post-test
1 (17) — (3) 5	1 (11) — (0) 5
1 (39) — (15) 8	1 (27) — (11) 8
2 (25) — (14) 7	2 (10) — (22) 7
5 (19) — (0) 8	5 (22) — (11) 8
4 (19) — (0) 8	

TABLE VIII

Mean percentage errors on Series III as a function of number of attributes differentiating stimulus figures

	Degree of difference between Standard and Comparison figures			
	One attribute			Two
	Orientation	Shape	Combined	attributes
Pre-test	18	11	13·3	6·7
Post-test	18·3	7·5	11·1	1·8

of confusions between figures differing only in shape and in both shape and orientation drops in the post-test. (Results for figures 3 and 6 have been omitted here.) This demonstrates an increasing refinement of discriminative ability with age; although errors are not eliminated, responses do reflect some appreciation of the aspects of similarity holding between the stimulus figures.

In terms of the proportions of correct responses, the order of difficulty of the three series is II, III, I, in both pre- and post-tests. This is to be expected, in view of the relatively greater difficulty of orientation compared with shape discrimination and of the fact that more criteria for differentiation are available in Series II than in Series III. What is not expected is the disparity in total numbers of responses; while these decrease for Series II and III between pre- and post-test, they increase for Series I. This increase, which is particularly noticeable in problems 3 and 4 and their repeats, appears to be linked to a change in distribution of responses which takes account of more global features of similarity such as obliquity. In problem 3, the figure in position 2 (position 5 in problem 3R), the only "vertical" figure in anotherwise "horizontal" array, is chosen 4 times out of a total for the two problems of 98 matches, and in problem 4, figures in positions 1 and 3 (3 and 2 respectively in 4R), the two non-oblique figures in the array, are picked 4 times out of 94 matches. The standards in these cases corresponded to the more often chosen class of comparison figures. These observations led us to hypothesize three stages in the development of form discrimination. The first stage, that of non-competence, would be characterized by a large number of errors and a small number of multiple responses (MRs), where more than one comparison figure is matched to the standard. This would be succeeded by an intermediate stage when the child has a clearer idea of what is required of him, but is not yet fully able to make the requisite unique match. Here we would expect to find fewer errors, and a concomitant increase in MRs; we would also expect to find that the figures given in the MRs would be more closely related to each other, in some at least intuitively reasonable manner, than to the remainder of the choice set. Finally, when a clear notion of the correct match is available we expect a large proportion of correct responses and very few MRs.

The appearance of this sequence, assuming it to be a valid representation of discrimination performance, will clearly be subject to individual differences in both subjects and problems; the level of responding will vary from child to child and from problem to problem. Nevertheless, we would expect increased success over the stages outlined above over time, with little if any regression to lower levels either with simpler problems at any one time, or with the same problems over time. That this very largely holds true for our experiment can be seen from Table IX. The second stage has been split to accommodate different types of performance which could not be assigned either to the first or third stages. The (arbitrary) criteria used to categorise our subjects are (first figures percentage MR, bracketed figures number of correct matches):

(A) $\leqslant 25$ ($\leqslant 25$)
(B) (i) $\leqslant 25$ ($> 25 \leqslant 75$)
(ii) $> 25 \leqslant 75$ ($> 25 \leqslant 75$)
(iii) > 25 (< 75)
(C) $\leqslant 25$ (> 75)

The sequence holds up reasonably well. Variations in individual performance and improvement are wide. Five subjects in Series I, 2 in Series II, and 1 in Series III show a poorer performance on the post-test than the pre-test, by these criteria; of these 8, 4 move from one stage to another, the rest staying within (B) but in a lower sub-division.

The relationship between choices based on position preference, as described for Series I, and figural similarity, remains to be clarified. If the sequence described above, particularly the intermediate stage, is to be accepted as valid, then we must show that responses by subjects classified at this level do reflect the latter rather than the former. We have already described the distribution of response to problems 3 and 4 and their repeats in terms of a tendency to ignore those comparison

TABLE IX

Classification of subjects as a function of their level of performance on all three series in pre- and post-tests

Stage		Series	I	II	III
(A)		Pre	5, 7, 8, 9, 12, 13,	8, 13, 15, 20	8, 13, 15, 20
		Post	14, 15, 20, 21, 22		
		Post	16, 18, 21	–	–
(B)	(i)	Pre	4, 6, 17, 18, 23	7, 12, 18, 21, 22, 23	7, 12, 18, 19, 21, 21, 22, 23
		Post	3, 4, 13, 14, 17	14, 16, 18, 21	6, 14, 16, 18, 21, 22
	(ii)	Pre	1, 10, 11, 16, 19	14	4, 14, 16, 17
		Post	6, 7, 8, 15, 19, 20, 21, 23	–	–
	(iii)	Pre	2, 3	3, 4, 10, 11, 16	1, 10, 11
		Post	10, 11, 12	5, 7, 8, 11, 15, 20	7, 8, 11, 15, 20
(C)		Pre	–	1, 2, 5, 6, 9, 17, 19	2, 3, 5, 6, 9
		Post	1, 2, 5, 9	1, 2, 3, 4, 6, 9, 10, 12, 13, 17, 19, 22, 23	1, 2, 3, 4, 5, 9, 10, 12, 13, 17, 19, 23

Numbers refer to children, assignment being arbitrary.

figures which are intuitively less closely related to the standard than are the rest of the array. In problens 4 and 4R one of these figures was always in position 3, the second most commonly chosen position in both the pre- and post-test. Table X shows the number of responses to these "obvious mismatch" (OM) figures in relation to the total number of matches given by subjects at stages B(ii) and B(iii) making MRs in these four problems. It is clear that these subjects are using criteria of global similarity in making their judgments, and that if position preference is operating it is certainly not dominant.

TABLE X

Influence of gross features of similarily on the responses of subjects in categories B(ii) and B(iii)

		No. subjects making MRs	Total responses (less "all")	No. "all" responses	Position of OM	No. responses to OM
Problem	3	17	43	2	2	1
	3R	13	30	2	5	2
	4	14	34	3	1, 3	1
	4R	10	28	1	3, 2	2

Discussion

The results of Series II and III show that even at three-and-a-half children are capable of discriminating between figures with considerable success, and that there are quite distinct class boundaries to the confusion errors which they do make. The shape of a figure, and whether or not it is open or closed, appear to be more often used than number of sides or orientation as criteria for differentiation. The number of attributes defining the difference between figures is also of importance, in that fewer confusions are made when there are two rather than one distinguishing features present; this is to be expected, inasmuch as there is a greater chance that the difference between two figures will be noticed. However, it is very probable that with a "messy" set of stimuli, such as that used in Series II, the counting of attributes becomes irrelevant (to the experimenter in his attempts to formalise what his subjects are doing) since some must be given a heavier weighing than others and because of the way in which attributes become inextricably confounded even in comparatively simple geometric figures. In both series the emphasis on shape at the older age, with orientation being very much less used as a determinant of similarity, suggests that whereas orientation can be used as a basis for judgement (from the information gained from Series I), where there is a more obvious criterion of differentiation, which shape may be inferred to be from the greater ease with which such problems are handled, that alone will be used; the child has no difficulties over the combination of attributes, because he is not allowing them equal weight. It may not be coincidental that it is also at this age that children tend to fail in the Piagetian conservation of liquid task through their inability to appreciate the interaction between two aspects of the situation, to realize that increased height is compensated for by decreased width.

As for orientation discrimination, once the child has passed the first stage in the developmental sequence outlined above, his responses are going to reflect, if not identity with the standard, at least some systematic relation of similarity between it and his choices. To this extent our results contradict a suggestion by Gibson *et al.* (1962), to the effect that transformation type is a better predictor of errors than the nature of the stimulus. Our findings are more in accord with Rosenblith's (1965) comment that figures may "behave" differently under similar transformations, that different error patterns may be obtained depending on the standards and the choice set available; it appears that the relationship between the features of the standards and of the members of the comparison array will largely determine the type of matches made, thus reflecting the importance of context in perceptual judgements.

In many ways the sequence of stages in the development of form discrimination which has been inferred from our results is the most interesting aspect of this study. The transition from the first to the second stage is analogous to a phenomenon noted, in a different context, by Donaldson (1963), an increase in response error following the partial development of new abilities in the child; opportunities for mistakes increase following the realization that more factors than had hitherto been considered must be taken into account, while the capacity to cope with these additional aspects of the situation is still under-developed. The problem here is in knowing what the developing abilities are. In the present context one of these would appear to be how similarity is understood and expressed. In the first stage a child who has an at least partially correct notion of what "same" means seems to be unable to apply it in any systematic fashion; matches to the standards are based more on proximity to it than on any other factor (witness the position effects, particularly in the pre-test, in Series, and the very large number of matches to figure 8 in both Series II and III). This is an effect which has also been noticed by Bijou and Baer (1963). Then in the second stage the child does start responding on the basis of similarity between stimulus figures, but without differentiating sufficiently between them to produce the correct matches of the final stage. However, there is a strong interaction effect, in that the level of responding appears to be at least partly a function of stimulus complexity. For example, in problems 1 and 2 of Series I the total number of responses dropped slightly in the post-test as compared with the pre-test, whereas it increased considerably for the rest of the problems. Given these findings, it seems reasonable to suggest that the greater number of errors children make in pair comparison situations is an artefact of the experimental set-up; there is no room here for partial success, so that unless the child has reached the third of the three stages described here he will be recorded as having "failed to make the discrimination".

A final implication of this experiment for the study of the development of form discrimination appears to us to be the importance of taking a relational point of view rather than an absolute one. The child's comprehension of the instructions, of the terms in which he is to compare the figures, should be independently checked and it is to be expected that the errors made will be a function not simply of the features of the stimuli taken in isolation but of the relationship between those of the standard and of the comparison figures. It then becomes more practicable to consider systematically changes in the use of contextually defined stimulus information as the child develops.

We should like to thank our many colleagues for helpful comments and the S.S.R.C. for research support.

References

BIJOU, S. W. and BAER, D. M. (1963). Some methodological contributions from a functional analysis of child development. In LIPSITT, L. P. and SPIKER, C. C. (Eds.), *Advances in Child Development and Behaviour*, Vol. 1. New York: Academic Press.

BIRCH, H. G. and LEFFORD, A. (1967). *Visual Differentation, Intersensory Integration and Voluntary Motor Control.* Lafayette, Society for Research in Child Development, Mon. Vol. 32, No. 2.

BRAINE, L. G. (1965). Age change in mode of perceiving geometric forms. *Psychonom.* **2**, 155–6,

BROWN, D. R. and Goldstein, J. A. (1967). Stimulus correlates of the discrimination behaviour of children. *Psychonom. Sci.* **9**, 177–8,

CAMPBELL, R. N. and YOUNG, B. M. (1968). Some aspects of classificatory behavour in pre-school children. (Presented to the Education Section of the British Psychological Society, September.)

DONALDSON, M. C. (1963). *A Study of Children's Thinking*. London: Tavistock.

DONALDSON, M. C. and WALES, R. J. (1968). On the acquisition of some relational terms. In HAYES, J. R. (Ed.), *Proceedings of the 1968 Carnegie Symposium on Cognition.* New York: Wiley.

DORAN, J. E. and HODSON, F. R. (1966). A digital computer analysis of palaeolithic flint assemblages. *Nature, Lond.* **210**, 688–9.

GARNER, W. R. (1966). To perceive is to know. *Psychol.* **21**, 11–9.

GHENT, L. (1960). Recognition by children of realistic figures presented in various orientations. *Can. J. Psychol.* 14, 249–56.

GHENT, L. and BERNSTEIN, L. (1961). Influence of orientation of geometric forms on their recognition by children. *Percept. Mot. Skills*, **12**, 95–101.

GIBSON, E. J. (1965). Learning to read. *Science, N.Y.* **148**, 1066–72.

GIBSON, E. J. GIBSON, J. J., PICK, A. and OSSER, H. (1962). A developmental study of the discrimination of letter-like forms. *J. comp. physiol. Psychol.* **55**, 897–906.

GIBSON, E. J., OSSER, H., SCHIFF. W. and SMITH, J. (1964). A basic research program on reading. Final Report on Co-operative Research Project No. 639 to the Office of Education, Department of Health, Education and Welfare.

HUTTENLOCHER, J. (1967*a*). Discrimination of figure orientation effects of relative position. *J. comp. physiol. Psychol.* **63**, 359–61.

HUTTENLOCHER, J. (1967*b*). Children's ability to order and orient objects. *Child Development*, **38**, 1169–76.

KRUSKAL, J. B. (1964*a*, *b*). Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, **29**, 1–27 and 115–29.

ROSENBLITH, J. F. (1965). Judgements of simple geometric forms by children. *Percept. Mot. Skills*, **21**, 947–90.

RUDEL, R. G. and TEUBER, H. L. (1963). Discrimination of direction of line in children. *J. comp. physiol. Psychol.* **56**, 892–8.

SEKULER, R. W. and HOULIHAN, K. (1968). Discrimination of mirror-images: choice time analysis of human adult performance. *Psychol.* **20**, 204–7.

SEKULER, R. W. and ROSENBLITH, J. F. (1964). Discrimination of direction of line and the effect of stimulus alignment. *Psychonom. Sci.* **1**, 143–4.

SHEPARD, R. N. (1962*a*, *b*). The analysis of proximities: multidimensional scaling with an unknown distance function, I and II. *Psychometrika*, **27**, 125–40 and 219–46.

WOHLWILL, J. F. and WIENER, M. (1964). Discrimination of form orientation in young children. *Child Development*, **35**, 1113–25.

Received 20 July 1970