A Developmental Study of Representation and Strategy in Children's Solutions to Problems Involving Chance and Probability.

Thesis submitted for the degree of Ph.D. at the University of Warwick by

Andrew Young,
Department of Education,
University of Warwick.
1974.

PAGE NUMBERS CUT OFF IN ORIGINAL

CONTENTS

CONTENTS (Cont'd.)

List of Figures.

List of Tables.

Facing page:

List of Appendices.

Acknowledgements

Help was received from many quarters during the preparation of this thesis, and for all of it I am extremely grateful. Special thanks are due to the staff and pupils of the Spon Gate J.M.I. school and Hearsall secondary school, Coventry, where most of the research was carried out. The headmasters of these schools, Mr. Coombe and Mr. Shilton, were most co-operative in allowing disruption of school routine at somewhat irregular intervals. Thanks are also due to the Research Unit of the University of Bristol School of Education for permission to use the Brimer cluster analysis, and to the staff of the Unit and the University's Computer Unit, especially Ann French and Chris Amos, who ran the program and helped sort out many technical hitches. Dr. J.G.Wallace supervised the research and contributed many of the ideas. Needless to say, this does not make him responsible for any misconceptions. In the preparation of the final draft much assistance was lent by Plym Peters, which was greatly appreciated. The research was supported by two post-graduate studentships from the Social Science Research Council.

Andrew Young

August, 1974.

ABSTRACT

The problems used in the study involve two collections of elements of two colours. The proportions of elements of each colour in each of the collections is varied, and the way children reason when asked which collection they would prefer in order to gamble for a specified outcome is investigated in three situations:

(a) The elements are beads to be drawn from boxes. (72 subjects aged 5-10 years, 48 subjects aged 11-14 years).

(b) The elements are single segments marked on circles of different sizes, with pointers to be spun. (72 subjects aged 6-11 years).

(c) The elements are similar to (b), but marked into separate pieces to allow comparison by counting. (60 subjects, aged 6-10 years).

Four possible ways of solving such problems are outlined:

Method 1: Guessing, alternating choices and other irrelevant methods.

Method 2: Comparing the amounts of the target elements in each collection, and choosing the collection with the greater amount.

Method 3: Comparing the differences between the amount of target and non-target elements in each collection, and choosing the collection with the most favourable difference.

Method 4: Comparing the proportions of target and non-target elements in each collection, and choosing the collection with the most favourable proportion.

Within the main age range investigated (6-10 years), methods 1-3 are found to form a developmental sequence in situation (a), whereas in situations (b) and (c) the predominant developmental sequence is from Method 1 to Method 2 only. It is argued that this can be explained by considering the methods of quantification used by subjects in each situation.

(A summary of the way in which the main themes are developed in the thesis is given at the end of the thesis.)

CHAPTER 1.

Introduction

One of the recurrent themes of cognitive psychology is that people are constantly trying to gain control of their environment by predicting what will happen and testing these predictions against outcomes. For example, Kelly (1963) says:

> 'Thus far we have said that the person is bent on anticipating events. His psychological processes are channelized with this in mind. Each person attunes his ear to the replicative themes he hears and each attunes his ear in a somewhat different way. But it is not mere certainty that man seeks; if that were so, he might take great delight in the repetitive ticking of the clock. More and more he seeks to anticipate all impending events of whatsoever nature. This means that he must develop a system in which the most unusual future can be anticipated in terms of a replicated aspect of the familiar past.' (Op.cit.,p.58)

Unfortunately many of the events we have to deal with are of a complexity which would make causal prediction impracticable for everyday use, and this has led to much research on 'subjective' or 'psychological' probability judgements in the last two decades. (Cohen and Christensen 1970; Peterson and Beach, 1967). The term 'subjective probability' is used to describe the probability of occurrence a person ascribes to an event rather than its mathematical or objective probability of occurrence, and the two need not be related. For example, the mathematical probability of winning by backing a particular number when playing roulette may be quite small, but the gambler may consider his chances of winning as much better than they are mathematically. He may even have reasons for doing so. The most celebrated example of such a reason is the Monte Carlo fallacy, which in fact depends upon a miscalculation of the mathematical probability of success, rather than a purely subjective estimate.

The easiest way of understanding the Monte Carlo fallacy is by means of an illustration. Suppose we toss a coin four times and each time it lands on heads. How likely is it that the next throw we make will give another head (accepting that no trickery of any kind is involved)? The chances of throwing five heads in a row are slim, in fact $\frac{1}{2}x\frac{1}{2}x\frac{1}{2}x\frac{1}{2}x\frac{1}{2}$, so surely the next throw is almost bound to be a tail?

To many people this is a very compelling line of reasoning, as is testified by many 'systems' for winning gambles. The reason it is fallacious is that every throw of the coin in independent of the throws which preceded it, so that the probability of heads coming up on any throw is always $\frac{1}{2}$. Admittedly the probability of throwing five heads in a row is $\frac{1}{2}x\frac{1}{2}x\frac{1}{2}x\frac{1}{2}x\frac{1}{2}$, but the probability of throwing four heads and then a tail on the fifth throw is also $\frac{1}{2}x\frac{1}{2}x\frac{1}{2}x\frac{1}{2}x\frac{1}{2}$. The point is that these probabilities only apply in the case when one is embarking on one's first throw; they are the probabilities of future events. In the example given the four heads have already been thrown and are therefore irrelevant to any future probability considerations.

Many other examples of factors which influence peoples judgements of probabilities have been collected by Cohen (1960, 1964) and Cohen and Christensen (1970). The flavour of this research is well illustrated by an experiment reported by Cohen (1957). In this experiment adult subjects were given a choice between different possible gambles in the form of lotteries. The choice was arranged to be between a single large probability or a number of smaller probabilities e.g. drawing either one ticket from a box of ten, or ten tickets from a box of one hundred, in the latter case putting back the ticket drawn each time before making the next draw. This makes the mathematical probability of success exactly the same in both cases, but the experimental subjects appear to have been more influenced by psychological than by mathematical considerations. About four-fifths of the subjects preferred to make the single draw from the box of ten, and even when as many as fifty draws from the box of one hundred were allowed, many subjects still preferred the single draw. Clearly, they were afraid that they would keep drawing the ticket they had just put back. This is confirmed by the fact that when the subjects were allowed to draw the ten tickets (or even fewer) from ten separate boxes, a majority of subjects changed to preference for the plural chance to the single draw from the box of ten tickets. In other words they changed from underestimating the plural chance to overestimating it. A similar experiment is reported by Cohen and Chesnick (1970).

The general conclusion that one can draw from Cohen's work is that peoples' estimates of the probabilities of events do not necessarily obey rules of mathematical probability, but are subject to different, psychological rules. However, this conclusion needs qualifying because it is only supported by empirical evidence in cases where calculation of the mathematical probability of a given event would be quite complex, so that it may be the case that estimates only obey psychological rules when people are unable to apply the mathematical rules.

If one accepts Cohen's thesis, with the qualification stated above, then it becomes pertinent to ask where these 'psychological rules' of probability come from, and how they are developed during childhood, adolescence, or whenever.

Some answers to these questions are provided in Cohen (1957), where studies are reported both of the ways in which children make judgements about the likelihood of chance events and the ways in which their understanding of probability terminology changes with age. In one experiment children were shown a display of two vertical columns of lights which were lit in succession, the childs' task being to guess whether the next light to be lit would be in the left or the right hand column. The children involved in the experiment were between six and eleven years old and the younger children seem to have followed a win-stay-lose-shift strategy. Older bhildren paid more attention to what was happening to the display and predicted that if more lights were lit on the left-hand side, the next light would appear on the right, and vice-versa.

Cohen considers that 'language of uncertainty' can be usefully thought of as having three main categories:

1. Words such as 'probability', 'surely', 'possibly', which denote a subjective probability and are potentially quantifiable.
2. Words like 'many', 'often', 'soon', which are also quantifiable, but denote not so much a condition of uncertainty as a quantity imprecisely known.
3. Words like 'fat', 'rich', 'drunk', which are not reducible to any accepted number because they are given various values by different people.

In his studies of the development of the meanings of such terms, Cohen has found that they are very context specific. For example, a child is asked to take 'some' sweets from a bowl and the experimenter notes how many he has taken. This number varies in different situations, such as if the child is alone or if there are other children with him who are also to take sweets, if there are a large or a small number of sweets available, and so on. It also depends on the items involved, 'some friends' and 'some trees' suggesting different numbers to most people. As well as this the meanings of such terms show a marked change with age. Children aged between six and fourteen years were told to take 'a few' or 'a lot' of beads from a tray and it was found that the older the child, the fewer beads he will take, and also that the difference between 'a lot' and 'a few' widens with increasing age. The expressions 'nearly always' and 'very rarely' show a similar development, increasing from about 2 to 1 for a child to 20 to 1 for a person twenty-five years old.

The general impression created by Cohen's work is one of considerable ingenuity in the design of individual experiments but a lack of any systematic approach to the area, and, above all, any coherent theoretical position. Cohen and Christensen (1970) excuse this on the following grounds:

'The danger of premature and excessive formalisation in psychology has been indicated with respect to the resulting impoverishment and distortion of the psychological actualities which the formal models are designed to represent. The get-rich-quick vogue, in the guise, for example, of over-hasty computerisation of psychological processes the complexities of which are unrecognised, is not likely to identify, let alone resolve, basic psychological issues'. (Op.cit.,p.142).

Although pleas for'open-mindedness' of this kind may well be useful in the opening stages of any research, once an opening has been made they are likely to prove malfunctional. Kuhn (1970) has put forward a convincing case for the priority of paradigms in the establishment of scientific traditions, and the way in which accepted paradigms determine both the kind of problems the practitioners of a science will attempt to resolve and their mode of resolution. The observation of any phenomenon and collection of data will always be guided by some sort of model, and it is desirable that this model should be made as explicit and open to inspection as possible. The 'distortion of reality' caused by our models of reality is something that must ultimately be lived with, not decried. This point is well made in the fundamental postulate of Kelly's personality theory, (Kelly, 1955) which is that a person's processes are psychologically channelized by the ways in which he anticipates events. Thus the best that we can do is to be continually modifying our models of the world in order to extend their scope and remove incompatibilites between our predictions and the observed outcomes.

A similar view of the nature of knowledge is taken by Piaget (1967). A clear exposition of Piaget's views is given by Furth (1969), who points out that knowing in Piaget's use is taken in a very general sense and does not imply any conscious or reflective knowing. Piaget holds that behaviour at all levels demonstrates aspects of structuring, and he identifies structuring with knowing. When something has been assimilated to the organism's structure, it is known, so that knowledge is not seen as a property of the independent object, nor is it in the subject, instead it is constructed by the subject as a relation between himself and the object. This is perhaps more easily understood if a clear example of assimilation functioning in this way is given:

'To illustrate briefly the concept of assimilation, take a baby who has acquired the ability to grasp things in his environment. Piaget conceptualizes this state of affairs by saying that the baby has a sensory-motor scheme of grasping. This grasping scheme functions by assimilating a great variety of external things to itself; in other words, the baby is observed to grasp and handle many different objects. These things have in common that they are amenable to grasping even though their specific figural outlines may differ one from the other. The grasping scheme corresponds to this common property of the objects or, even better according to Piaget's theory, confers this common property on them. (Furth 1969, p.14).

The following quotation is also very useful, as a concise summary of Piaget's position concerning knowledge:

'......... knowledge is in Piaget's theory never a state, whether subjective, representative, or objective. It is an activity. It can be viewed as a structuring of the environment according to underlying subjective structures or as a structuring of the subject in living interaction with the environment. In any case, the laws of structuring are seen as intrinsically related to the self-regulations which are found at all levels of a developing biological organisation.' (Furth, 1969 p. 20-21.)

Such consideration of the nature of knowledge and scientific discovery makes Cohen and Christensen's concern to avoid impoverishing and distorting psychological 'actualities' by 'premature and excessive formalisation' appear somewhat simplistic. As has already been pointed out, such concerns may have a place in the opening stages of a field of research, but significant advances will only be possible when theoretical models and assumptions have been made explicit.

In this thesis an attempt will be made to take a more systematic look at a narrowed down area in this field, namely the way in which children make judgements about chance events in which all the mathematically relevant information is available to inspection, though not necessarily explicit. Before doing this, however, it is necessary to draw attention to certain distinctions being made here which are not always made.

Firstly, it is necessary to distinguish events which we normally see as chance events, such as the throwing of dice, coins, and so on, from events to which we can ascribe probabilities, but do not normally see as 'due to' chance, such as whether it will rain tomorrow. In the literature all events to which probabilities can be given are often lumped together, particularly when subjective probability is used as part of a theory of decision making. This may be a valid step in certain contexts but for present purposes it will be better to maintain this distinction until evidence that it is unnecessary is produced.

The other distinction to be made is between chance events where all the information needed to deduce the mathematical probability of success is readily available, and those where it isn't. In the former category would be throwing dice, coins and so on, in the latter would be lotteries where one doesn't know how many people have bought tickets, and the kind of situations usually involved in probability-learning experiments. The same distinctions are drawn by Goffman (1967).

To reiterate, this thesis will concentrate on the way children make judgements about events which are commonly regarded as 'pure chance' but in which all the mathematically relevant information is available in some form.

Within this narrowed down field the main contribution has undoubtedly been made by Piaget and Inhelder (1951). The point of departure of Piaget's work on the development of childrens' conceptions of probability is well stated by Flavell (1963):

> 'In order to identify a set of phenomena as 'chance events' one has first to identify a set of phenomena which are not chance events, a non-chance ground against which chance can emerge as a figure. Only if cognitive processes are developed enough to order and organize the intrinsically certain, lawful, and predictable by means of rational operations, can things which are intrisically uncertain, unlawful, and unpredictable be apprehended as such' (Op.cit pp. 341-342).

In order to investigate what he calls the child's' conception of probability' Piaget devised several experiments. These will not all be reported here as they are susceptible to the usual criticisms levelled at Piaget's experiments, lack of experimental controls, detail in reporting, and adequate tests of the significance of the results. However, in cases where other researchers have followed Piaget's lead his originall experiment will be considered.

One example of this is the work of Yost, Siegel and Andrews (1962). They argue, as does Braine (1959,1962), that Piaget may be underestimating his subjects' intellectual capacities by relying on verbal responses, and repeated one of his experiments in a form which required non-verbal responses and gave concrete rewards for correct responses. As carried out by Piaget and Inhelder (1951) the experiment involved two collections of counters. There were two kinds of counter in each collection and the child knew how many counters of each kind were in each collection. His task was then to judge whether he had a better chance of drawing a counter of a certain kind from one collection or the other when the counters were face down. The modifications made to this experiment by

Yost et al (1962) involved using a different kind of display, demanding less understanding of probabilistic terminology, controlling for irrelevant colour preferences, accepting non-verbal responses (i.e. choices without reasons) and reinforcing 'correct' responses. Their criterion of success was correct choice (i.e. choice of the gamble which is not likely to lead to success on mathematical grounds) rather than the 'correct choice and correct reason' favoured by Piaget. According to Yost et al (1962) a higher number of correct responses in 24 trials was produced by four-year olds with the non-verbal than the Piagetian experimental technique. They deduce from this that four-year olds do have some understanding of probability. Unfortunately this conclusion is invalid because their experiment confounds true understanding with the kind of learning involved in probability-learning experiments.

The work has been followed up by Carlson (1970). He compared the performance of children rather older than those in the Yost study on a task similar to Yost's and a Piaget derived task. His Piaget task involved a red ball being propelled by a spring. The distances it travelled in ten trials were marked and the child was asked how far it would go the next time. The child had to give the right answer, a right reason, and resist a counter-suggestion by the experimenter. From the results of this experiment Carlson concludes:

(a) The development of probability reasoning increases monotonically with age.

(b) Verbal and non-verbal assessment techniques do not measure the same aspects of this development.

(c) The general age brackets suggested by Piaget for the ontogenesis of probabilistic thinking are supported.

(d) Sex and intelligence are not significant variables in the development of probabilistic thought.

Of these conclusions perhaps (c) is most interesting, that is the experiment confirms Piaget's reported findings. (Carlson (1969), reports other experiments which also confirm Piaget's findings.)

Conclusion (a), that the development of probability reasoning increases monotonically with age, can only be tentatively drawn from evidence gathered from two problems. Conclusion (d) that sex and intelligence are not significant variables in the development of probabilistic thought is supported by Carlson (1969), but is strictly not supported by Carlson and McMillan (1970), which shows that mental retardation does affect children's probability abilities. Conclusion (b) that verbal and non-verbal techniques do not measure the same aspects of the development of probabilistic thinking is not surprising in view of the difference between the stringent criterion of success employed in the Piagetian experiment and the lax criterion of the non-verbal experiment (in which a subject has 0.5 probability of guessing correctly). A difference between non-verbal and Piagetian measures of children's probability concepts was also found by Davies (1965) using children between three and nine years of age, and by Goldberg (1966) with pre-school children.

The experiments mentioned so far support the view that the Piagetian and non-verbal techniques for investigating children's probability concepts may not be tapping the same abilities. They also lead to greater confidence in Piaget's results as they appear to be easily replicable. But replication of results does not reveal anything more about what the children are actually doing in these experiments,and what kind of knowledge their probability knowledge is. The only coherent theoretical perspective on this appears to belong to Piaget. Having stated that chance events can only be understood in contrast to non-chance events, he claims that this is a characteristic development during the concrete operational period. According to Piaget the pre-operational child is unable to differentiate chance and non-chance events, but when he begins to apply operations to different phenomena he can discover areas where they fail to give definite knowledge. This allows the differentiation of chance and non-chance events, and with further development in operational thinking it is possible for the child to establish that certain events are more likely to occur than others.

There are two achievements which Piaget sees as pre-requisites for a full understanding of probability. These are:

(a) The ability to work out all possible outcomes of events.

(b) The ability to handle proportions.

Both of these achievements are characteristic of the formal operational period and presuppose the development of what Piaget calls a combinatorial system. (See Piaget and Inhelder, 1966, for a useful summary of this). The child in the concrete operational period can only reason about things he can see to be true, and cannot deal with hypotheses. The essence of formal operational thought is that hypotheses can be systematically generated and assessed, and this permits the attainment of full understanding of probability. In fact Piaget believes that the notion of probability arises from assimilation of the concept of chance to the formal operations.

The trouble with this kind of theory, as Bruner (1966) has pointed out, is that it is more like a logical description of what a psychological theory must account for than a psychological theory. However, it provides a useful starting point, and makes Cohen's position seem ad-hoc by comparison.

Preliminary Experiment

In order to ascertain which lines of inquiry in this field might prove most fruitful, a small pilot experiment was run. This involved among others a problem devised by simplifying some of Cohen's experiments to a level suitable for children and a problem designed to throw some light on Piaget's claim for the necessity of a 'combinatorial system' which develops during the formal operational period.

Method: 17 subjects aged between 5 and 17 years were tested and one aged over 21 years. Five problems were involved in the experiment, and a complete list can be found in Appendix A. The two problems which are particularly relevant to the present discussion are as follows:

Problem 1: This is a simple form of the kind of problem used by Cohen.

The problem involves a forced choice situation in which the subject can choose to throw a die to get a six or toss a coin to get a head. He is asked the reason for his choice and given a 'smartie' as reward if his gamble has a successful outcome. The problem is extended by adding extra throws with the die, i.e.

1(b) two throws of the die to get a six or one toss of the coin to get a head.

1(c) three throws of the die to get a six or one toss of the coin to get a head.

1(d) four throws of the die to get a six or one toss of the coin to get a head.

1(e) five throws of the die to get a six or one toss of the coin to get a head.

Problem 4: It was hoped that this might elucidate Piaget's claims concerning the 'combinatorial system'.

4(a) The subject has to say whether the outcome of one toss of a coin will be a head or a tail, giving a reason for his answer. The coin is then tossed, and success rewarded with a 'smartie'.

4(b) This is the same as 4(a) except that the subject has to predict whether the outcome of two tosses of one coin will be two heads, two tails, or one head and one tail. (The alternatives are not necessarily offered in this order).

4(c) This is the same as 4(b), except that the subject is asked to predict whether the outcome of three tosses will be three heads, three tails, two heads and a tail, or two tails and a head.

The order in which the five experimental problems were presented to each subject was varied, but the subparts of each problem were always given in the same order. All subjects were rewarded with a 'smartie' whenever a choice led to a successful outcome.

Results: The criterion of success used was the choice of the mathematically most likely outcome, or an assertion to the effect that one could choose either way where both choices had the same mathematical probability of success, together with what the experimenter regarded as an adequate justification. Full results of the experiment and sample responses can be found in Appendix A.

The results of the experiment showed a steady improvement in performance with increasing age. A product-moment correlation showed the correlation of performance with age to be nearly 0.8., which is significant at the 0.001 level. (See Appendix A). More interestingly the results showed some of the 'psychological rules' to which Cohen refers. All subjects of over 13 years thought that three throws of the die to get a six had the same chance of success as one toss of a coin to get a head. The reasoning behind this appears to be:

Each throw of the die has one-sixth probability of success.
There are three throws.
Hence the probability of success on three throws is one-sixth $x\ 3 = \frac{1}{2}$.
But the probability of success with one toss of the coin is also $\frac{1}{2}$.
So it doesn't matter which we choose.

This is intuitively convincing but mathematically false. It is false because there is an implicit assumption that one cannot throw the same number more than one time in the three throws, whereas in fact this might happen.

In this situation subjects clearly did not use a combinatorial system to deduce all the possible outcomes and then compare the number of favourable with the number of possible instances in order to calculate the probability. Arguably, the initial estimates of one-sixth and a half for the one-trial probabilities of success with die and coin might have been arrived at in this way, but once this was done a simple mathematical formula was employed. In fact, use of a combinatorial system would have been extremely arduous. The results of problem 4 showed a high proportion of correct responses at all ages tested. (The younger children were not tested on part (c)). However, the answers given for parts (b) (predicting the outcome of two tosses of a coin) and (c) (predicting the outcome of three tosses of a coin) seemed to be mainly 'intuitive', in that subjects were unable to formulate clear reasons for their choices.

Part (a) (predicting the outcome of one toss of a coin) was successfully performed (correct choice and correct reason) by nearly all subjects over age 9. This is an age at which children do not possess the combinatorial system, according to Piaget. This means that either the knowledge that heads or tails is equally likely to 'come up' when one throws a coin cannot be dependent on such a system, or the system develops at an earlier age than Piaget believes. There was also little evidence of the use of a combinatorial system in the results of parts (b) and (c). The few successful answers could be ascribed to use of such a system, but it appears to be linked more closely to the subject's mathematical sophistication than anything else. Pire (1958) reached a similar conclusion after analysing a large number of results to similar problems.

This experiment was intended as a guideline for further work and is clearly inconclusive because of its modest scale. However, there are some possible criticisms of this kind of experiment which are worth considering now. The first of these criticisms is that the effect of reinforcement in the form of experimenter's approval or material reward is not allowed for. The consistency of the children who have reasons for their choices in sticking to them, even when disappointed, rules this out. Of course the children who are 'just guessing' may well be influenced by pay-off considerations, but this doesn't affect the results when reasons as well as choices are analysed. Other possible criticisms are that the criterion of success employed of correct choice and correct reason is too stringent, and that one shouldn't make so much use of verbal data anyway.

These criticisms have been lumped together because they make a similar point, namely that this kind of experiment underestimates the abilities of any given age group of children. Such criticisms are misdirected because the experiment is not intended to generate normative data, and as long as one accepts this one might as well use the richest data source available, namely verbal data. There will always be children who make correct choices consistently and yet can't explain why. Indeed it is commonly accepted that competence of this kind generally precedes linguistic performance, but this is only likely to be the case when competence is newly acquired.

The general conclusion to be derived from this little pilot study is that whilst Piaget's idea of a combinatorial system may provide an adequate logical description of the way probabilities are correctly calculated in this kind of situation, it is not a very plausible psychological theory of what people actually do in such situations. Cohen's claims about 'psychological rules' being employed to generate subjective probabilities offer more promise, but look somewhat empty in view of the ad hoc way in which these psychological rules must be inferred from situation to situation.

Introduction to the type of problem to be investigated.

What seems to be needed is a detailed study of what children actually do in order to make judgements of probabilities. This must involve either acceptance of verbal data or else an exceptionally high degreee of methodological rigour in testing alternative explanations of the childrens' performances. Many of the studies reported so far fail to do either of these things, serving merely as replications of Piaget's observations. However, in some cases Piaget's observations are detailed enough to be worthy of further consideration.

One example of this is the problem, already discussed, which was criticised by Yost et al (1962). The experimenter makes up two collections of counters, some of which have a cross on one side and some of which don't. For example, one collection might contain two counters with crosses and two without crosses, whilst the other collection has two counters with crosses and one without a cross. The collections are then turned face down and their members are scrambled up. The child is asked whether he has a greater chance of drawing a counter with a cross from one collection or the other. After investigating children's answers to a number of problems of this kind, using varying proportions of crosses and non-crosses, Piaget reported the following developmental stages:

(a) The child does not apply any systematic and relevant strategy.

(b) The child predicts on the basis of the absolute number of counters with crosses in each collection, rather than the ratio of these to the total number of counters.

(c) The child predicts by comparing the proportions of crosses and non-crosses in each collection.

Stages (b) and (c) are confirmed by Lowe and Ranyard (1973) in a study involving children aged 8 to 11, who should be in the later part of Piaget's concrete operational period. Stage (b) is described by Lowe and Ranyard as involving use of magnitude cues, and stage (c) involves use of proportion cues. Information about the cues used by the the subjects was obtained by designing the problems so that proportion and magnitude cues were in conflict and then carrying out individual regression analyses to determine which strategy (magnitude or proportion) matched the subject's judgements more closely. These analyses revealed that most children quite distinctly utilized one cue rather than the other. However, the children who did use proportion as a cue were not in the formal operations stage, although Lowe and Ranyard think that they may have been 'helped' by certain details of the design of their experiments. One of their experiments involved collections of red and green beads in beakers, about which they comment:

> 'The results in the Beaker test seem in direct conflict with Piaget and Inhelder's findings. This may be due to the way the information was arranged on our slides, compared to the way Piaget and Inhelder set out the information (in the same task). On our slides, the favourable and unfavourable cases were arranged in separate columns in each beaker emphasizing this comparison, whereas Piaget and Inhelder set out each collection randomly. When the information is organized appropriately many children will be capable of comparing serial orders of favourable and unfavourable cases and basing their judgements on a simple compensation between these. If the collections are arranged unsystematically such a strategy will not occur spontaneously and attempts to make the more difficult comparison between favourable and possible cases are more likely.' (Op.cit.,p.7).

Later on they say:

> 'In conclusion, we are suggesting that when information is organized appropriately, children at the concrete operations stage can use a successful proportion - like strategy in a probability task.' (Op.cit., p.8).

The fact that this particular type of problem has been so popular with investigators renders it suitable for more detailed analysis, which will be attempted in this thesis.

Introduction to the theoretical approach to be adopted

Having surveyed some of the researches in the field chosen, the problem arises of what to do next. The resolution of this problem depends very much on the individuals' conception of what science is and should be like. The positivist point of view is that laws should be produced from the generalisation of experimental data and organised into a formal theoretical structure. Attempts to introduce theoretical concepts with a standing of their own are frowned upon as metaphysical in cases where the entities they refer to are not susceptible of direct observation.

Harré (1971) has argued that this approach to science is empty and restricting. He believes that the way forward in any science is an adequate conception of the nature of the entity being studied, so that chemistry, for example, really got underway with the idea that materials were made up of minute parts. He goes on to say:

> 'Conceptions of the nature of human beings up to now have been either insufficiently specific, for example as with the Skinnerian conception of man, which does not allow, on the basis of that conception alone, people to be distinguished from pigeons (Skinner, 1953), or insufficiently powerful, as, for example, the early Millerian conception of a manifold of information channels which makes the capacity to ignore, reflect upon, and control output inexplicable (Miller, 1964, pp 171-184)'. (Op.cit., p. 115.)

Harre suggests that the concept of a person should become the basis of the new post-positivist psychology. The traditional conception of a person includes the idea of rational agency and the idea that people are capable not only of monitoring and controlling their performances, but of monitoring the control they excerise in the first order performance.

> 'To put this point in another useful way, not only do people follow rules, but they know that they follow rules and they may, and often do, choose the rules they will follow in accordance with rules for choosing rules.' (Harré,1971,p.116)

This conception of man as a person is undoubtedly powerful, especially in accounting for peoples' social behaviour,with which Harré seems mainly concerned, but it may itself be lacking in specificity when applied to the more traditional topics of the psychology of cognition. However, the quotation given above connects directly with another, more precise conception of man, the conception of man as not only a person, but a person who is continually processing information. Such a conception,of course,lies behind most modern cognitive psychology and an early form of it was rejected by Harré in one of the passages quoted. However, the criticism given by Harré can only be fairly applied to Miller's early work stemming from information theory, and is much less applicable to later developments of the information processing approach. These developments have become particularly associated with the school of thought dominated by Newell and Simon. (Newell and Simon, 1961, 1972).

Reitman (1965) provides an account of the information processing approach to psychological problems. He explains that if each person is considered as made up of receptors, effectors and a control system joining them, the information processing approach involves concentrating on the control system and avoiding most of the questions concerned with sensory and motor activities. The type of control system usually postulated in order to explain the content and direction of thought comprises a number of memories containing symbolised information which are interconnected by various ordering relations, a number of primitive information processes which operate on the information in the memories, and a well-defined set of rules for combining these primitive processes into whole program of processing.

The aim of an information processing theory is to build up, from these basic parts, first a flow chart then a computer programs which will produce the same behaviour as a person in a particular situation. This program is then accorded the status of a theory of the persons psychological processes and is tested in the same ways as any other psychological theory. This in an extension of the famous insights of Turing (1950).

It is important to note that the essence of the information processing approach is the description of postulated internal mechanisms in an abstract but precise manner. Hence no direct correspondences between these internal mechanisms and physiological mechanisms need be specified or known, since the theory stands or falls only by how well it accounts for what people do when thinking. This means that any information processing theory might be implemented in a variety of ways, and much interest has centred on the possibility of computer simulation which this introduces.

There are various advantages to be gained from producing information processing theories in the form of computer programs. The main advantage is that this provides a very rigorous test of the theory concerned. With models of other kinds it is quite possible to overlook ambiguities and errors, and the more complex the theory is the more this is likely to happen. The conclusions reached by running a computer program follow with certainty from the processes built into the program, and in addition the sequence of processes performed by the computer can be used as an additional source of data to be compared with human protocols obtained in the equivalent situation.

Despite the undoubted validity of these arguments, there are certain practical considerations which must be taken into account before throwing in one's lot with the simulation school. One of the most important of these is that the level at which simulation is possible depends on the richness of the data available. The point is neatly put by Miller, Galanter and Pribram, (1960):

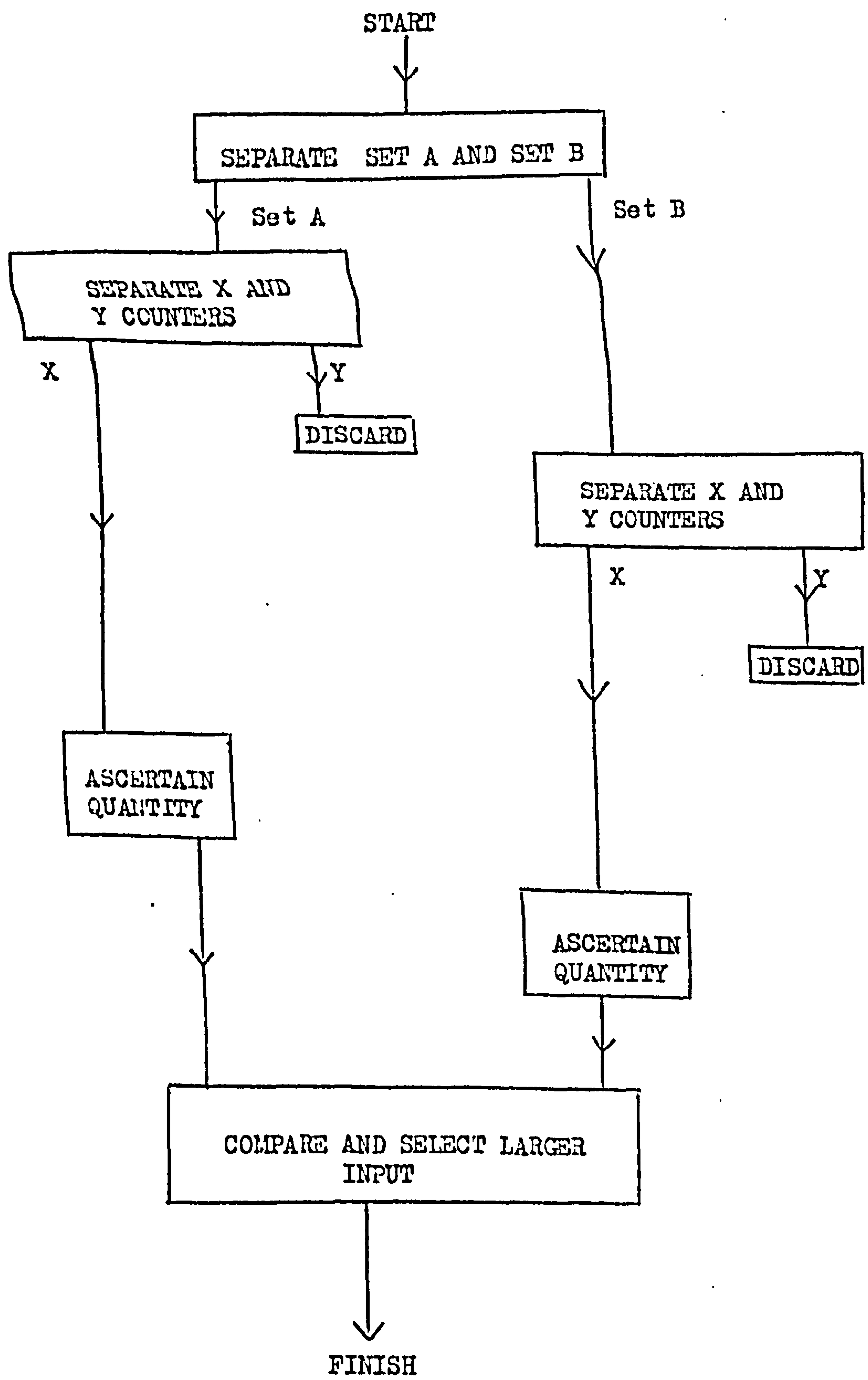
> 'Another thing that Turing's theorem did - or should have done - was to focus attention on the adequacy of the description of behaviour. A machine cannot be expected to simulate something that has never been described - it can be held responsible only for those aspects of behavior that an observer has recorded. No simulation is complete and no simulation preserves all the characteristics of behavior.' (Op.cit.,p.47).

In view of this the approach adopted here will be to utilise as rich a source of data as is practicable and to construct models of psychological processes, without introducing the amount of detail which would be needed to construct an equivalent computer program. Because of this the models will be not only less detailed than computer programs, but also capable of less complexity in practice, as the symbolic manipulations involved in interpreting complex models seem to present people with great difficulties. In this respect the proposed models will initially be more like what are often referred to as strategies (e.g. Bruner, Goodnow and Austin, 1956), but the aim is to refine them to a point where computer simulation might be possible.

The initial test of any theoretical model must be,as Newell and Simon (1972) maintain, whether it is sufficient to produce the data it is meant to explain. However, whether the model is sufficient in this way depends very much on the level of detail at which the sufficiency test is carried out. Again the aim must be to begin with a model which is sufficient to generate the gross aspects of performance and then modify it so that increasinly detailed behaviours can be predicted. Gregg and Simon (1967) also emphasize universality, precision, simplicity and flexibility as criteria on which to assess the usefulness of different models when there are no clear-cut differences in empirical predictions. These criteria are derived from Popper (1959).

So far the information-processing approach has been discussed in a general context of human cognition, and nothing has been said about how it can be applied to the type of questions associated with developmental psychology. One way in which this might be done is described by Klahr and Wallace (1970a);

FIGURE 1: MODEL 2: A POSSIBLE MODEL CORRESPONDING TO THE SECOND DEVELOPMENTAL STAGE REPORTED BY PIAGET AND INHELDER (1951).



(X is always used to refer to target counters.)

'We believe that the major task facing the child who has just been presented with an experimental task is to assemble from his repertoire of fundamental information-handling processes, a routine that is sufficient to pass the task at hand. We view the information processing demands of the task as being analogous to the compilation and execution of a computer program.......... Incoming visual and verbal stimuli are first encoded into internal representations. Then the assembly system attempts to construct, from its repertoire of fundamental processes, a task-specific routine that is sufficient to meet the demands of the verbal instructions. Having assembled such a routine, the system then executes it.' (Op.cit., p.362.)

The different stages children seem to go through in their ways of solving various tasks can then be described by different information processing models. In order to evaluate such models Klahr and Wallace (1970b) propose the formal criterion of developmental tractability as an addition to the criteria specified by Gregg and Simon (1967). Developmental tractability is used to describe the ease with which a model can be interpreted as both a predecessor and successor of other models in a developmental sequence. By this criterion the best model is the one which is most amenable to transformation into a model of a later developmental stage, or which is itself most easily seen as a transformation of an earlier model. The purpose of using such a criterion is to permit assembly of a collection of stage models which will facilitate inferences regarding the nature of the transition mechanism responsible for the stage to stage development.

Some possible models.

An attempt will now be made to apply the research strategy outlined above to the results of the other researches which have been reviewed. The first stage described by Piaget, when the child does not seem to apply any systematic and relevant strategy to solve the problem, looks like it is a collection of different behaviours with little in common except their irrelevance by adult standards. Because of this it will be ignored at present. The best place to begin the analysis thus becomes Piaget's second stage. This is the stage where the child predicts on the basis of the absolute number of the target counters in each collection, rather than the ratio of these to the total number of counters. If the collections of counters are called Set A and Set B, the counters the child is trying to get (the target counters) X, and the non-target counters Y, then one possible model corresponding to Piaget's second stage is shown in figure 1.

FIGURE 2: MODEL 3: A POSSIBLE IMPROVEMENT TO MODEL 2 (FIGURE 1).

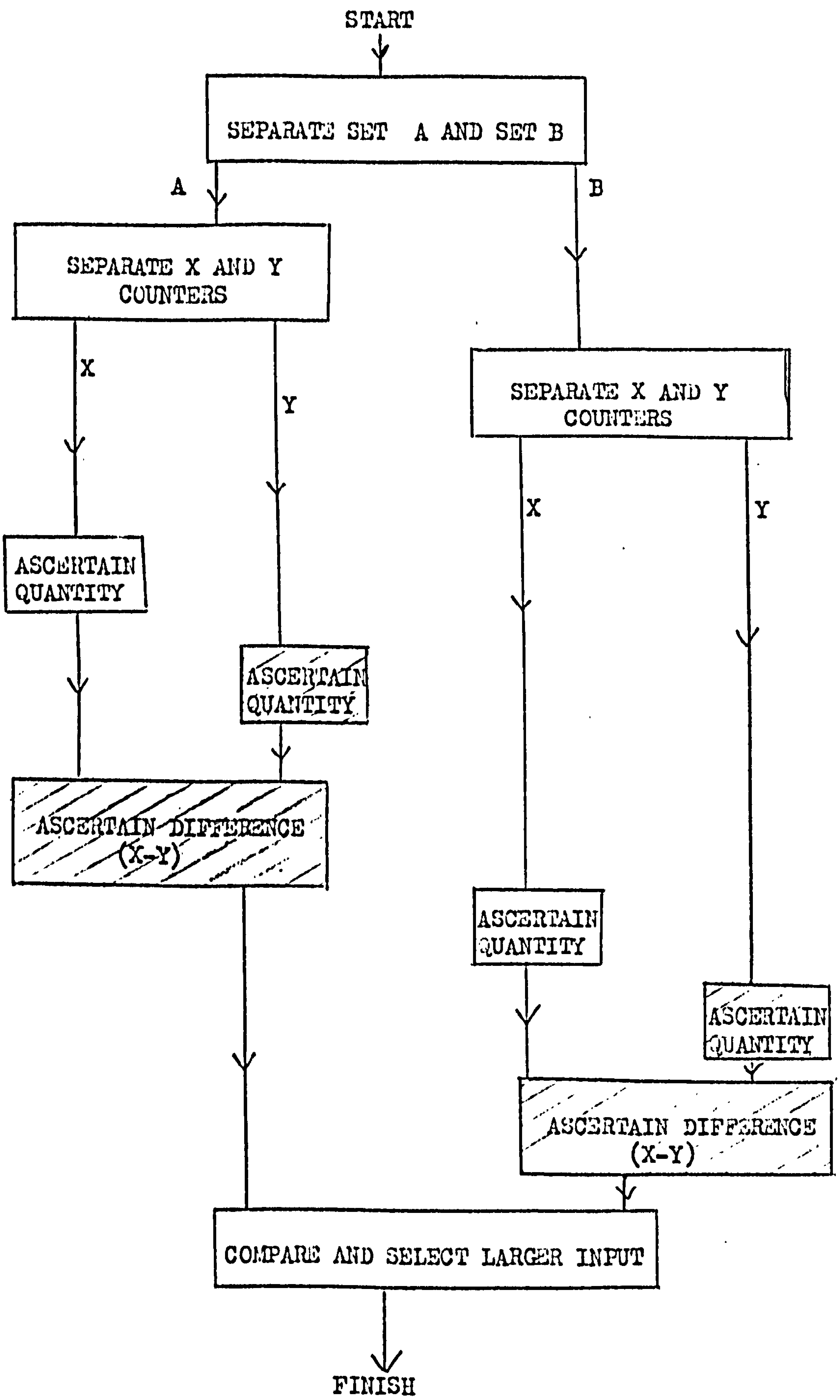


Figure 1 is not intended as a conventional flow diagram, it is merely a convenient way of representing the strategy so that people can understand what it involves. In other words, although it is not in the normal form of a flow diagram, it states all the necessary processing to achieve a solution, without specifying how such processing might be implemented. The advantage of doing this is that it doesn't commit the modeller to any one implementation and yet is precise enough to lead to testable predictions. There is no commitment, for example, to a parallel or sequential processor since although the model is set out in a parallel form (when read from top to bottom), it could equally well be considered sequentially by reading it from left to right and top to bottom.

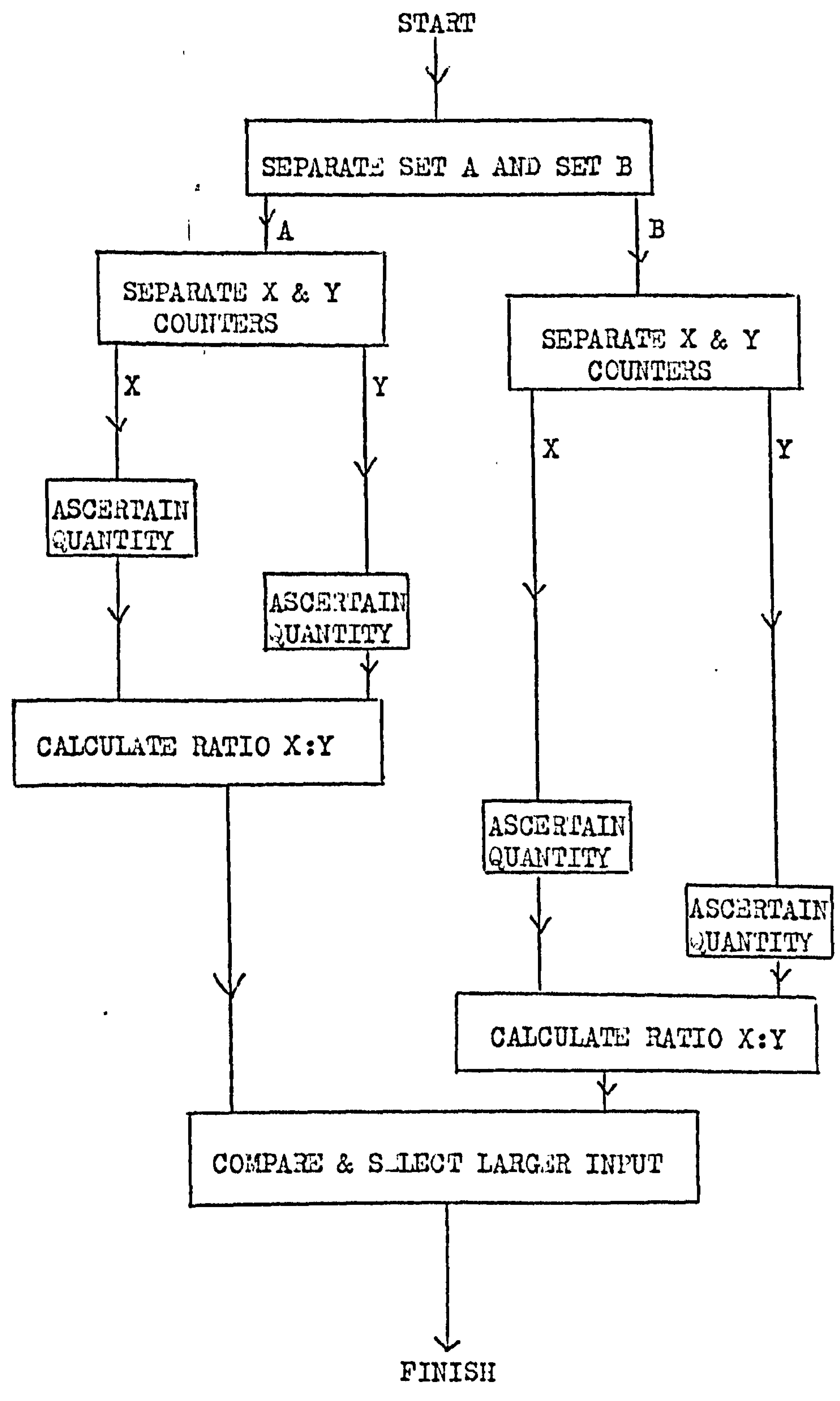
The model generates the same answers as Piaget's second stage child in most situations. However, whenever the number of X counters in each collection is the same, it will be unable to produce an answer. Fortunately there are two simple modifications that can be made to the model in order to get an answer out of it. It could be instructed to choose at random whenever there is no unique outcome, or it could be made to take account of the number of Y counters in each collection (by repeating the strategy on the Y counters and choosing the collection with less).

These possible revisions will be referred to as model 2(a) and model 2 (b), respectively. Both model 2(a) and model 2(b) can be regarded as models of the second stage described by Piaget. It would be interesting to know whether children always reason like model 2(a) or like model 2(b) in the special situation where there are the same number of X counters in each collection, or whether some children reason like model 2(a) and some like model 2(b), or if neither model applies. Unfortunately the research reviewed so far cannot answer this question, so that it will have to be resolved by experiment.

The main limiting factor to model2's ability to produce mathematically satisfactory solutions is clearly its almost total disregard of information concerning Y counters. In fact model 2(a) makes no use at all of such information, and model 2(b) only uses it in the special case where Ax and Bx are equal. The simplest way in which information about Y could be utilised would be to make choices based on the difference between the number of X and Y counters in each collection. A diagram to represent this would be figure 2 (which will be called model 3).

The shaded portions of model 3 show how the original model (model 2) would have to be modified to take this form. This shows

FIGURE 3: MODEL 4: A 'TRUE PROPORTION' MODEL.



clearly the relation between the previous model and this elaborated version, but it is somewhat oversimple. This is because the final comparison (made in the bottom box) may be quite complex. The differences to be compared may be positive or negative, and there are various ways in which this problem might be handled. A model 3(i) might be proposed which would have a number system capable of dealing with negatives, or a model 3(ii) in which various transformation rules are applied to make sure that the final comparison is always between positive numbers. However, there is a simple reason why little attention need be paid to such possibilities at the moment. The reason is that the difference between models 3(i) and 3(ii) is not as great as the difference between model 2(a) and model 2(b), in the sense that one couldn't devise a simple test to distinguish which one is operating. The main empirical difference appears to be the kind of mistake each is likely to make. If operating correctly, however, both models 3(i) and 3(ii) will produce the same results so that for present purposes they will not be distinguished and will be referred to as model 3. In other words, model 3 is being used to describe the general strategy of comparing (Ax - Ay) with (Bx - By) and need not specify the precise manner in which this strategy is carried out. The strategy is surprisingly powerful, and leads to mathematically correct answers to many problems of this type.

At this point it would be pertinent to ask why so much space is being devoted to what is, after all, only a possible way of handling the problem and is not reported by Piaget as one of the developmental stages children go through. The answer to this question can perhaps be more easily understood by first considering what a 'true proportion' model, which would correspond to Piaget's third stage and would generate mathematically correct answers, might entail. The general layout would have to be something like figure 3.

Now this is almost the same as model 3, except that where model 3 compares X and Y by substraction, this uses division. This means that the proposed model compares the ratio of favourable to unfavourable cases instead of the ratio of favourable to possible cases (which is the mathematical concept of probability). This is quite legitimate, as the research reported by Fischbein, Pampu and Manzat (1970) demonstrates that children generally prefer to make the favourable to unfavourable rather than the favourable to possible comparison. Of course this represents a very considerable increase over model 3 in the complexity of the processing being done, but the sequence of processing steps remains essentially the same.

Aims for the first main experiment.

Model 3 thus serves as a theoretical link between model 2 and the 'true proportion' model 4. Both model 2 and model 4 can be seen as models of the stages described by Piaget (his second and third stages respectively) but model 3 is not described by him. However, it is so much easier to convert model 2 into model 3, and then model 3 into model 4, instead of making the huge jump from model 2 to model 4 directly, that to anyone considering the child as being in some ways analogous to a developing information processing system it is puzzling that model 3 doesn't occur.

One possible explanation is that model 3 does in fact occur, but has been overlooked by most previous researchers. Support for this explanation can be derived from the point already mentioned that model 3 is a powerful way of handling problems of this type, and very often produces the same choice as model 4. This means that unless the researcher was specifically looking for model 3, or had generated by accident problems which would distinguish model 3 and model 4, answers reached by model 3 could easily be mistaken for the results of 'proportional' reasoning. This could be what Lowe and Ranyard (1973) mean by the following remarks:

> 'When the infromation is organised appropriately many children will be capable of comparing serial orders of favourable and unfavourable cases and basing their judgements on a simple compensation between these.' (Op.cit., p.7.)
>
> 'Estimating the proportion of red, using the non-verbal response mode, could be based on a comparison of the number of red with the number of green dots using a set of quite simple concrete operations.' (Op.cit.,p.8).

The latter remark certainly indicates that a similar strategy to model 3 may be used by children when estimating proportions. If this is the case why shouldn't it also be used when comparing probabilities? To the writer's knowledge the only study where the model 3 strategy is explicitly acknowledged as a possible way of comparing probabilities is that of Fischbein et al (1970). They find that although the strategy is used by some subjects none of them used it in solving twelve or more out of the eighteen problems involved in their experiment. However, their experiment involved children of rather disparate age groups: pre-schoolers (5.0 to 6.4), third-graders (aged 9.0 to 10.0), and sixth graders (aged 12.4 to 13.7). This means that it is possible that the strategy may be more important at some of the ages not sampled. Such considerations can only be satisfactorily resolved by empirical evidence. The experiment proposed has the following aims:

A. To find out whether the models put forward provide satisfactory models of childrens' reasoning when faced with this type of problem.

B. To find out whether model 2(a) or model 2(b) provides the more accurate description of Piaget's second developmental stage, or whether both apply.

C. To find out whether model 3 is only a theoretical possibility, or if it corresponds to an actual strategy employed by children to solve problems of this type.

D. To find out whether there is a developmental sequence from the first stage described by Piaget to model 2 to model 3 to model 4, or even a sequence from Piaget's first stage to model 2(a) to model 2(b) to model 3 to model 4.

These aims are not, of course, mutually exclusive. The problems proposed will be similar to the kind already described. However, instead of collections of face-down counters, collections of red and green beads in boxes will be used. The problem then becomes one of choosing one of two boxes to draw a red or green bead from. (These materials are similar to those used by Yost et al, 1962). Fuller details of this will be given later.

Methodological considerations: data collection and analysis.

There are two main kinds of data which can be obtained in an experiment of the proposed type, namely the choice made by the subject and the reason he gives for his choice. The first source is not particularly convenient because the choice to be made is only between two alternatives, so that the subject can guess correctly on half the trials. This means that any experiment which uses correct choice as its success criterion will havecto involve a large number of trials and rely heavily on a statistical estimate of the 'significance' of the results. Of course one could always increase the number of alternatives the subject could choose from in order to reduce the required number of trials, but this would involve an undesirable increase in the complexity of the problem. The danger here is that when the problem becomes too complex for the children to handle without excessive difficulty, they will revert to 'guessing'. In addition any experiment which is analysed solely in terms of the choices made by the subjects will need to be exceptionally well designed because of the ease with which alternative explanations of these choices can be adduced. The pitfalls inherent in this approach are well demonstrated by the study by Yost, Siegel and McMichael (1961), similar to one already described (Yost et al, 1962),which tries to remove the verbal element from the presentation as far as possible, as well as from the analysis. Yost, Siegel and McMichael reach the conclusion that their experiment demonstrates that four-year olds do have some understanding of probability, in contrast to Piaget's view that children under seven years are unable to respond consistently to the quantitative proportions of the elements involved. However, as Flavell (1963) points out,

> 'It is uncertain......... whether this method succeeded in 'liberating' pre-existing concepts and strategies by virtue of its procedural differences from the other method, or whether it was simply a more effective training procedure for inculcating response patterns (and perhaps concepts as well) which the child did not have in his repertoire when he walked into the experimental room'. (Op.cit., p.393.)

This criticism is not supported by Goldberg (1966). She compared results to an experiment of this kind elicited by a method similar to Piaget's techniques with those elicited by a method similar to that of Yost. She argues that if Flavell's interpretation of Yost's result, that the methodology used provides a better learning condition than Piaget's methodology and not a better technique for assessing available concepts, is correct, then subjects' performance should improve more during a number of Yost-type trials than during a number of Piaget-type trials. What she found was small but significant learning effects in both situations, which she claims does not support Flavell's interpretation. However, performance on the Yost-type trials was better than that on the Piaget-type trials, and it is arguable that the kind of learning which went on in the different types of trial might be of a different nature. The point being made is not that this did or did not happen, but that it is remarkably easy to produce alternative explanations of this type, so that as long as only non-verbal data is being used our knowledge is not advancing very fast. Furthermore, the argument has become fixed on a point which is of little importance anyway, for all that Yost and his colleagues have done is to show that if one makes a certain kind of problem simple enough for pre-school children to grasp, then they seem to show some ability in solving it. There is nothing very mysterious about this, unless one holds that 'abilities' do not develop but appear suddenly at fixed chronological ages. How the pre-school children solve the problems, what they are actually doing, is not revealed by experiments of this type.

The second possible source of data is verbal data, i.e., the reason the child gives for his performance. Reliance on verbal data and on verbal methodology has been criticised by Braine (1959,1962). Braine points out that by using assessment techniques which do not distinguish intellectual capacities from associated verbal abilities, the intellectual capacities of children are likely to be underestimated. This criticism, however, is only important in cases where normative claims are being made, so that it is not strictly relevant to the present project.

Brainerd (1973a) claims that the judgement criterion is superior to explanation for establishing the presence of a cognitive structure because it allows safer inferences concerning correspondences and sequences between hypothesised structures. However, his argument is not completely convincing (see Reese and Schack, 1974;. Brainerd,1974) and he admits himself that there is one case in particular where explanations are most useful:

> 'Explanations can supplement judgements in such a way that one is provided with insights into the nature of the structure or structures under consideration.' (Brainerd, 1973a,p.178).

This means that for present purposes explanations may be of great value.

Another claim advanced by Braine (1959,1962) is that it must be impossible to study the development of a concept by using methods which employ verbalccues to evoke the concept. Put simply, the argument runs that if the child understands the verbal cue, he must have already developed the concept. Hence if he responds appropriately we know for certain that he has the concept, but if he doesn't respond appropriately we have learnt very little. In addition it is often maintained that childrens' comprehension of linguistic items precedes their production of those items (e.g. Fraser, Bellugi and Brown,1963) so that reasons given for decisions may well lag behind the 'real' reasons.

Nelson (1973), on the other hand, argues that production may be 'instrumental' in advancing comprehension:

> 'As strategies of language learning, comprehension and production are both complementary and contrastive. A child who produces little may be relying on internal processing of the language he hears to advance his linguistic competence. On the other hand, production may be instrumental in advancing comprehension. The relation between these factors is not at

all well understood at the present time.' (Op.cit., p.40).

The final resolution of these problems cannot be achieved until the function of language in cognitive development and the relationship between linguistic and cognitive development is properly understood. At present this is one of the most controversial areas of research and any adequate review would be beyond the possible scope of this study. Piaget and Inhelder (1966), and Sinclair-deZwart (1969) seem to view language development as generally subordinate to cognitive development, although they do admit the facilitating effect of a well-developed language system on cognitive development. Much greater stress is placed on language as an organising factor in cognitive development by Luria and Yudovich (1959), whilst intermediate positions are taken by Vygotsky (1962), and Bruner, Olver and Greenfield (1966).

The confused state of this area means that no firm decision can be made as to the advisability or inadvisability of utilising verbal data. However, it has already been asserted that the exclusive use of non-verbal data is a desperate measure because of the difficulty involved in designing a watertight experiment and the paucity of such data. This means that it will be worthwhile at least to consider ways in which verbalisations might be tentatively returned to play.

The importance of this task, from the standpoint of the philosophy of science, is strongly expressed by Harré (1971), who argues that the accounts people give to themselves as well as to others in explanation and justification of their actions:

> '......... are not just pieces of 'verbal behaviour' produced by controlling variables; these are systems of statements in whose content and organisation lies the missing dimension of psychology. Far from inventing all sorts of ingenious ways of trying to find out about people by not asking them about their experiences, the Copernican revolution in psychology consists exactly in getting them to give as much as possible of anticipatory, monitoring and retrospective commentary upon their actions, i.e., to give accounts.' (Op.cit., p.116, underlined passages italicised in original).

One very obvious step to take is to make the instructions and procedures of any experiment as concrete as possible, so that illustrations are provided by the experimenter at every step of what he means by whatever he might be saying at that time. In addition the use of any but the simplest words should be avoided, so that hopefully the combination of simple instructions and concrete examples may obviate some of the possible criticisms advanced by

Braine concerning lack of understanding by the child of what is required of him, and such like.

The probelm of how to analyse verbal results, in the form of reasons given by the child to justify his choices, is more difficult. Such reasons are often very sketchy and make it tempting for the experimenter to interpret their 'real' meaning. The method adopted by Piaget, of picking out 'typical' verbalisations corresponding to the different stages postulated, is unsatisfactory because of its subjective and preconceived nature. There is also the possibility that any reason given will be a post hoc justification of a choice made because of some totally different considerations, answers like; 'I like the colour', or 'My Mummy told me to', can easily be regarded as of this type, while even answers like 'It's the bigger one' are imprecise enough to be consistent with an alarming variety of interpretations.

In terms of current psychological theory, as already mentioned, this problem appears intractable without an accepted model of the language - thought interrelationship and its development. In practice, a workable position may be reached if the following conditions are satisfied:

(a) Design condition: The experimental design is arranged to 'disambiguate' the reasons for their choices given by the subjects as much as possible.

(b) Results condition: The choices made by the subjects and their justifications of these choices do not conflict with each other.

(c) Analysis condition: The analysis of the verbal data involves a minimal amount of interpretation by the experimenter.

There are of course other necessary conditions, such as that the experimenter does not suggest possible reasons to the subject, which apply to all experiments and are commonly accepted.

The results condition, (b), is one whose fulfilment can only be observed empirically. If choices made and reasons given are consistent, then it is worth pursuing the verbal analysis, if the choices and reasons frequently conflict only non-verbal data can be validly used. How far the design condition, (a), can be met depends mainly on the individual experimenter's ingenuity, but it will obviously be helpful to have a clear idea of what kind of reasons the subjects are likely to produce in the experiment and possible explanations of these reasons.

The analysis condition, (c), is perhaps the most important and the most equivocal. In some ways it is linked to (a), because if the results of an experiment are unambiguous then less 'interpretation' will be needed than if they are not. Bryant (1971a) pointed out how easily alternative explanations of Piaget's research findings can be devised and has tried to produce more tightly-controlled versions of some of the standard Genevan situations (Bryant, 1971b, Bryant and Trabasso, 1971). The dangers implicit in attempts to infer cognitive structures from childrens' verbalisations have long been recognised by Piaget (e.g. Piaget, 1932), but as Wallace (1972a) remarks:

> '......There is little evidence in his practice that he has coped successfully with them. In particular, he appears to have failed to resist the supreme temptation of capitalizing on the ambiguity of verbal responses to derive support for his preconceptions.......there is a case for viewing the childrens' verbalisations cited as simply illustrations of the appropriateness of a preconceived theory.' (Op.cit.p.23)

The problem to be faced, then, is that of devising a rigorous way of encoding verbal data for analytic purposes which will involve minimal interpretation of those data by the experimenter. Inter-experimenter consistency of interpretation, such as might be obtained by correlating the results of analyses by different experimenters, is not by itself sufficient if the experimenters have previously been trained in the use of the scoring system, as this is equivalent to training them to make the same interpretations (Cooper, Costello, Douglas, Ingleby and Turner, 1974). The need for such a technique has been stressed by Dienes (1959) and Smedslund (1964).

One technique for the classification of qualitative data which looks promisingly rigorous and objective has been devised by Brimer (1967, 1973). The only interpretation made by the experimenter in the Brimer analysis is whether any pair of responses generated in the experiment are the same or different. In this way categories of indistinguishable responses to the same item can be built up and used as the basis of a cluster analysis which will show which subjects produce response categories that are statistically associated. The experimenter's personal interpretation of his own results can then be compared with this more objective analysis to see whether it is confirmed or falsified. The actual procedures involved in the Brimer analysis will be treated at greater length elsewhere, for present purposes it is sufficient to say that this method of analysis satisfies the requirement (c) which was proposed as a condition for making use of verbal responses, and so it will be used as the principal method of analysis of experimental results.

CHAPTER 2.

The First Beads Experiment

Having decided that verbal response data are acceptable providing certain conditions are met, it is now necessary to return to the detailed design of the experimental problems. These can be arranged to distinguish the various models of what children might do which have been advanced. For example, in order to distinguish use of a method of solution equivalent to model 2 (a) from use of a method equivalent to model 2(b), a problem is required which has the same number of target beads in each collection but a different number of non-target beads.

e.g.	Collection A	Collection B	
	XXX	XXX	X = Red bead
	YYYY	YY	Y = Green bead

When the target colour is red (X) there are an equal number of red beads in each collection.

Hence: model 2(a) → choose either collection
model 2(b) → quantify green and choose the collection with less green, i.e. B.

Thus, in situations of this type, model 2(b) will always lead to a choice of B, whereas model 2(a) can lead to a choice of either A or B. In addition model 2(a) implies that it doesn't matter which choice is made, so that if a child says something to this effect it can be regarded as compelling evidence in favour of model 2(a), as he is contradicting the 'set' implicit in the experimental design, which suggests that there is a reason for favouring one collection over the other. Of course if the child is intimidated by the experimental setting, or if model 2(a) is not a very 'strong tendency' then he may simply choose A or B without saying that he thinks it 'doesn't really matter'. In some cases there may even be a regression to a reason of the type described by Piaget as the first developmental stage, such as 'It's pretty', or 'It's lucky'. It must also be admitted that model 2(a) and model 2(b) can only be distinguished in this way if answers consistent with the model 2 type of reasoning rather than models 3 or 4, or Piagets first stage (for which no model has been advanced but which can conveniently be referred to as model 1) have been obtained consistently in response to other problems. Otherwise models 3 and 4 produce similar choices (although not necessarily similar reasons) to model 2 (b), and model 1 and model 2(a) also lead to similar patterns of choices.

To distinguish model 2 and model 3 or 4 a problem of the following kind may be used:

Collection A	Collection B	
XXXX	XX	X = Red bead
YYYYYY	YY	Y = Green bead

When the target colour is red (X).

Model 2 → Choose A because there are more reds in A than B.

Model 3 → Choose B because A has more greens than reds.

Model 4 → Choose B because of the unfavourable ratio of greens to reds in A.

Choices corresponding to model 1 in this type of problem will still be erratic, but model 2 will consistently select A, and models 3 or 4 will consistently go for B. Model 3 and model 4 may then be distinguished by problems of this type:

Collection A	Collection B
XXXX	XX
YYYYYY	YYYY

In this case, model 2 → choose A, whether the target colour is red or green, as there areemore reds and more greens in A than in B. Model 3 → choose either, whether the target is red or green, as there are two more greens than reds in each collection.Model 4 → choose B when the target is green, A when the target is red.

Again, a child reasoning by a method equivalent to model 3 will contradict the experimental set by asserting it doesn't matter whether he chooses A or B. The possibility of regression to model 2 or even advancement to model 4 when forced to choose also exists. If a significant number of subjects do claim that choice of A or B is equally likely to lead to success, strong evidence that model 3 represents one of the strategies children use to solve problems of this kind has been provided.

Another type of problem which will distinguish model 3 and model 4 is:

Collection A	Collection B
XXXX	XX
YYYYYYYY	YYYY

Here model 2 → choose A whether the target colour is red or green, as there are more reds and more greens in A than in B.

Model 3 → choose A when the target is green, as there are four more greens than reds compared with only two more in B. Choose B when the target is red, as there are only two less reds than greens and there are four less in A.

Model 4 → choose either, whether the target is red or green, as there are twice as many greens as reds in each collection.

In this case, then,model 4 is the one which will result in contradiction of the experimental set.

By using carefully designed problems of the types illustrated the applicability of the various models to the data could be assessed solely from a combination of choice-data and regression analyses. A similar method was used by Lowe and Ranyard (1973), and though there is much to recommend it, it suffers from the disadvantage of requiring each child to perform a large number of trials. This disadvantage would be very severe in this case, where several models are to be tested (Lowe and Ranyard were only testing between two models), although the work could be broken down into a number of smaller experiments designed to test pairs of models. However, a case has been made for the introduction of verbal responses (in the form of justifications made by the children for their choices) into the analysis, and if this is done a lot of information can be derived from less specific problems. For example:

Collection A	Collection B
XXX	XXXX
YYYY	YY

If the target colour is red(X):

A justification consistent with model 1 (Piaget's first developmental stage) would be: 'I'll have the one on the left/ right because its pretty'.

A justification consistent with model 2 would be 'I'll have the one on the right because there's more reds than there are on the left'.

A justification consistent with model 3 would be 'I'll have the one on the right because there's more reds than greens and the other has less reds than greens'.

A justification consistent with model 4 would be 'I'll choose the one on the right because there's twice as many reds as greens and the other has four greens onto three reds'.

As long as the child's answers are as explicit as those outlined, and there is no contradiction between the justifications given and the choices made, they can be used with some confidence as indicators of the way he reasons. Unfortunately problems will arise when unclear or ambiguous answers are given;

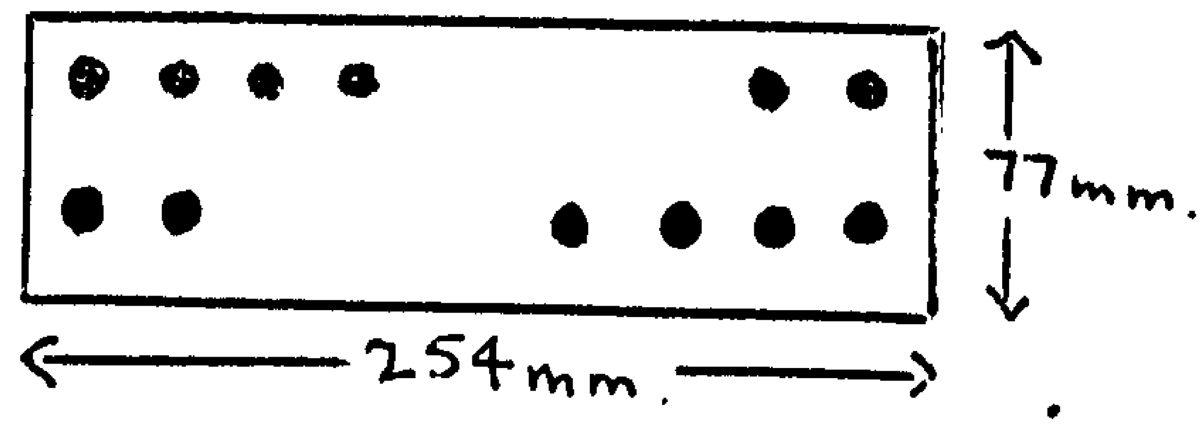
e.g. 'I'll choose the one on the right because there's more reds'. In this context 'more reds' might mean 'more reds than on the left' (model 2) or 'more reds than greens' (model 3). Some kind of further questioning would then be required to disambiguate this response. The experimenter might ask 'What do you mean by that?,' or 'Show me by pointing what you mean by that.' Failing this he could even ask 'How many more are there?' An answer to this question generated by model 2 would be that there is one more, because there are four reds in collection B and three reds in collection A. An answer generated by model 3 would be that there are two more, because there are four reds and only two greens in collection B. However, such specific 'prodding' as this is undesirable as it suggests to the child that quantification is relevant to the solution of the problem and it causes variations in the treatments each child receives. The question 'How many more are there?' is also not as straightforward as it appears, a point which will be elaborated later.

The choice to be made, then, lies between the conciseness of the experiment and its discriminatory power. An experiment involving few trials will have to rely heavily on verbal response data, whilst one involving large numbers of trials will suffer from problems associated with inattentiveness of subjects and the like, which are most important when working with children. After some pilot work, which indicated that the approach being advocated here is feasible, an experiment was designed which it was hoped represented a satisfactory compromise. This will be referred to as the first beads experiment.

<u>The first beads experiment</u>

<u>Procedure</u>: The aims of the experiment have already been outlined and will not be repeated. The experiment is in three parts, beginning with a pretest which is intended to allow the subjects to become familiar with the experimental situation, materials and instructions. This is followed by the main experimental problems, which are presented in an order varied from child to child. After this there are some more specific problems which are only given to certain children in certain circumstances and are intended as a more rigorous substitute for the 'clinical' interviewing technique often employed at the end of Genevan experiments. Throughout the experiment data are collected by the experimenter in the form of a written record of the reasons given by the subject to justify his choices, the choices he makes, and their success or failure in producing the desired outcomes.

The materials used comprise two boxes, each with a hole in the top large enough for one bead at a time to pass through. For each problem there are cards which have two collections of red and green spots marked on them; indicating how many beads of each colour are in each box. These arrays of spots are arranged in two ranks in one to one correspondence as far as possible, so that the card for Problem B1 would look like this:



The spots are 13mm diameter.

The subject places one red bead on each of the red spots and one green bead on each of the green spots. The left hand collection of beads is then transferred into the left hand box and shaken, and the right hand collection is transferred to the right hand box and shaken. The card representing the two collections remains in front of the subject, and he is asked which box he would choose to draw a bead out of if a red (or green as the case may be) bead is wanted, and the reason for his choice. If this reason is unclear he may be asked 'What do you mean by that?' The verbatim instructions are as follows for the case when red is the target colour:

'We're going to try out some games, and I want you to tell me what you think of them.'

What we do is we put beads on these spots. (Show appropriate card.) We put a red bead on each red spot, like this (demonstrate), and a green bead on each green spot, like this (demonstrate), Then, when we've done that these beads go in this box (point) and these beads go in this box (point). Can you do that for me?

(When this hasbbeen done). Now shake up the boxes so that you don't know where they are inside.

(When this has been done). You don't know where they are inside now, do you?

But you do know that these beads (point to collection A on display) went in this box (point to box A) and these beads (point to collection B on display) went in this box (point to box B).

Good. Now listen carefully and I'll tell you what we're going to do. We're going to tip just one of the boxes and let one bead out, and I want the bead that comes out to be a red bead. So which box would you tip?'

Now why did you choose that box rather than the other one? (If the reply to this is not clear, ask 'What do you mean by that?') When an answer has been given,

'Well, we'll try tipping that box and if you're right you get a point, if you're wrong you get nothing. If a red bead comes out you get a point, if its a green one you get nothing, O.K.?'

'You have to tip it very gently or they all come out at once.'

These instructions seem very complicated but turned out to be quite easy for children to grasp, probably because of the very direct way they relate to the materials used. Any difficulties experienced were overcome by repetition of the appropriate part of the instructions, but this was only rarely necessary. After the initial tril a simplified version of the instructions was used. Each problem is in two parts, namely the part where red is the target colour and the part where green is the target colour, so that bead drawn out on the previous trial would be returned to its box and the next instruction is:

'Now this time we'regoing to tip one of the boxes and let one bead out, and I want the bead that comes out to be a green bead. So which box would you tip?

Now why did you choose that box rather than the other one?' (If the reply to this is not clear, ask 'What do you mean by that?')

When an answer has been given 'Well, we'll try tipping that box and if you're right you get a point, if you're wrong you get nothing. If a green bead comes out you get a point, if its a red one you get nothing. O.K.?'

With each new problem it was also necessary to empty the boxes, put the beads on the new problem card, and transfer them to the boxes. (All this was done by the subject.) The instructions could then be repeated in the shortened form indicated. The instructions were not read out, but were memorised by the experimenter. This must have led to some inaccuracies caused by slips of the tongue, lapses of memory and suchlike, but it avoided the rather artifical nature of a 'read out' instruction. In general every effort was made to keep the experimental setting as relaxed and informal as possible. For this reason tape and video recording equipment was not used since if the experimenter writes things down this is far less intimidating than if the child's performance is recorded. On the other hand effort was also made to avoid giving any kind of specific encouragement or nonstandard instruction in order to minimise the effects of experimenter bias and non-identical treatments. The decision not to use video or tape means that the more sopisticated non-verbal measures, such as latencies and hesitations, are not available for analysis.

<u>Sample:</u> The subjects who took part in the experiment were seventy-two children aged between five and ten years from a local primary school, and forty eight children aged between eleven and fourteen years from a local secondary school. These were divided into groups of twelve children aged five, six, seven, eight, nine, ten, eleven, twelve, thirteen and fourteen years. As well as this four children failed

to finish the experiment and were replaced. These were two six year olds and a five year old, who seemed too intimidated by the experimental situation to want to finish, and an eight year old described by his teacher as having emotional problems. None of the subjects taking part in the experiment,or any of the subsequent experiments, was red-green colour blind. The catchment area of both schools was the same, being a working-class suburb of Coventry, but apart from this no attempt was made to control for socio-economic differences, intelligence differences, or sex differences. The primary school was of modern 'open design' and children were selected either by approaching them and asking them to take part or because they approached the experimenter at a time when it was possible for them to be fitted in. In the secondary school selection was done by class teachers who were asked to send 'ordinary' children. The somewhat casual nature of this sampling procedure can be excused on the grounds that only the possible strategies used by the children are being studied, and any relation these strategies may show to chronological age will be qualified by saying that it is only the case for the sample used. Conclusions about the possible relation of any strategies found to mental age, sex, or social class variables will not be drawn.

Problems used: Two problems were used in the pretest, whose results were recorded but not analysed, and which was intended solely to counteract any artifacts which might otherwise be introduced into the experiment, through lack of familiarity with the situation and materials. The pretest problems are:

Pretest A.	3G 2R	V	2G 3R
i.e.	YYY		YY
	XX		XXX
Pretest B	3G 4R	V	5G 4R
i.e.	YYY		YYYYY
	XXXX		XXXX

After the pretest there were four problems which constitute the main part of the experiment. These were presented in a haphazard order determined by shuffling the four cards corresponding to the four problems. This is easier to put into effect than a 'systematic randomisation' involving some sort of Latin square and should provide sufficient control against order effects and suchlike. In addition the order of red and green being first target colour with each new problem was varied either by alternating this order between problems or by doing two problems with one colour as first target and then the remaining two problems with the other colour as first target. Position preferences were controlled by presenting the cards either way up, so that the left-right order of the collections as shown below was often reversed. The experimental problems are listed below, with a short rationale for each one.

Problem B1:	4G 2R	v	2G 4R	
i.e.	YYYY		YY	Y = green bead
	XX		XXXX	X - red bead

This is intended as a 'simple' problem which could allow children to perform slightly better than they might do with more complex problems. It is desirable for the experimental problems to vary in level of 'difficulty' in this way, so that the flexibility of the strategies children use might be investigated. Mixtures of model 2 and model 3 responses might be obtained from the same child, whereas if the problems were of even difficulty these might be all model 2 or model 3 responses.

Problem B2:	4G 3R	v	2G 3R
i.e.	YYYY		YY
	XXX		XXX

For children who give answers predominantly of the model 2 type on the other problems, this will provide a test between model 2(a) and model 2(b). It is also possible that such children might show regression to the model 1 type of answer, or improvement to model 3 or 4.

Problem B3:	4G 4R	v	2G 2R
i.e.	YYYY		YY
	XXXX		XX

This represents a case where model 2 and model 3 or 4 responses can be clearly distinguished. Model 2 will lead to choice of the left-hand box in the both parts of the problem, whereas model 3 and model 4 allow choice of either box in both cases. The problem also represents something of a paradox, because model 2 leads to the conclusion that the left-hand box is better for drawing both green and red. It will be interesting to see if children notice this, and if it will affect their decisions. It is possible that some children will not be willing to choose the same box in both parts of the problem, and this could lead to changes of reasons in the form of improvements, regressions, or outright guesses.

Problem B4:	6G 4R	v	2G 2R
i.e.	YYYYYY		YY
	XXXX		XX

Again, model 2 can be distinguished from model 3 or 4. Model 2 would lead to choice of the left-hand box, whether red or green is the target colour, model 3 and model 4 will lead to choice of the left-hand box for green and the right-hand box for red. The main purpose of the problem, however, is to make some assessment of how much children's focus of attention is determined by the strategies they are using. Although model 2 implies that one should choose the left-hand box when the target colour is red, because there are more red beads in it than in the right hand box, it is also apparent that there are more green beads in it than there are red beads. This additional information may be discounted, or it may lead to some attempt to modify the strategy to assimilate it. The paradox that model 2 predicts that the left-hand box is 'better' for red and green, which was pointed out in problem B3, is also present here and may also contributed to some sort of change in strategy.

These four problems, and the two pretests, constituted the standardised part of the experiment. They were followed by three additional problems which were given only if the child's performance in the main experiment had met certain conditions.

The first additional problem is problem B5,

Problem B5:	8G 4R	v	2G 2R
i.e.	YYYYYYYY		YY
	XXXX		XX

This is essentially the same as problem B4 but in an extreme form. It was given to children who had produced two model 2 type answers to problem B4, but had possibly shown some signs of 'wavering', and also to any other children whose answers might have been interesting to the experimenter.

Problems B6 and problem B7 are designed to distinguish answers produced by reasoning equivalent to model 3 from those of model 2 or model 4. These problems were both given to children who had shown any sign of the model 3 type of answer.

Problem B6:	6G 4R	v	4G 2R
i.e.	YYYYYY		YYYY
	XXXX		XX

In this problem model 2 leads to a choice of the left-hand box for red or green, model 3 implies that either box will do for either colour (because there are two more greens than reds in each), and model 4 leads to a choice of the right-hand box for green and the left-hand box for red. There are also possibilitites that subjects whose previous answers have been predominantly of a type attributable to model 3 may

Table 1: Problems used in the first beads experiment.

Pretest. A. 3G2R v 2G3R

i.e. YYY YY (X = red bead, Y = green bead).

XX XXX

B. 3G4R v 5G4R

i.e. YYY YYYYY

XXXX XXXX

Study Problems.

B1. 4G2R v 2G4R

i.e. YYYY YY

XX XXXX

B2. 4G3R v 2G3R

i.e. YYYY YY

XXX XXX

B3. 4G4R v 2G2R.

i.e. YYYY YY

XXXX XX

B4. 6G4R v 2G2R.

YYYYYY YY

XXXX XX

Additional problems:

B5. 8G4R v 2G2R

YYYYYYYY YY

XXXX XX

B6. 6G4R v 4G2R

YYYYYY YYYY

XXXX XX

B7. 8G4R v 4G2R

YYYYYYYY YYYY

XXXX XX

The experiment consists of pretests A and B, followed by problems B1, B2, B3, B4, in randomised order, then B5, B6, B7, at the experimenter's discretion.

'advance' to model 4 or 'regress' to model 2 when faced with the necessity of choosing one box or the other. Regressions may be particularly common when red is the target colour, because there are 'only' two reds in the right-hand box.

Problem B7:	8G 4R	v	4G 2R
i.e.	YYYYYYYY		YYYY
	XXXX		XX

Model 2 leads to a left-hand box choice for red or green, but model 3 leads to choice of the left-hand box when green is the target colour (there are four more greens than reds and only two more in the other box) and the right-hand box when red is the target colour (there are only two less reds than greens and four less in the otherbbox). This time model 4 implies that either box may be chosen for either colour, as there are equal proportions of red and green in each box. The problem is also arranged to use quite simple proportions (1 : 2 in each case). In spite of this some regression from model 4 to model 3 answers is a possibility.

The problemsused in the first beads experiment have been collected together in Table 1.

Results of the first beads experiment.

The data collected consisted of the choices made by the subjects and the reasons they gave for their choices. The reasons given were written down at the time, and consequently are not authentic to the degree a tape recording might be. Often words like 'the' and 'there's' were missed out in haste,but any numbers and relational terms were specially noted. The order of presentation of the problems was also recorded, in case any investigation of 'order effects' proved necessary, but these data were not used in the main analysis. Any verbalisations other than choice reasons produced by the subjects were recorded at the experimenter's discretion, but were also ignored in the main analysis, as were the outcomes of the gambles.

The analysis to be carried out involves three parts:

1. An analysis of the type of reasoning shown by the children and the applicability of the theoretical models in terms of interpretations of the results made by the experimenter in accordance with a set of evaluative criteria.
2. An analysis of the results by means of the Brimer cluster analysis.
3. A comparison of the experimenter's analysis with the results of the cluster analysis.

Attempts to establish reliability between different markers in the first part of the analysis were not made because it was considered that training people to make the same interpretations is not a big step towards objectivity. Obviously such inter-scorer reliability would be an essential pre-requisite of any adequate test of children's knowledge concerning probability matters, but this is not the purpose of the experiment. It has been argued that for the present purposes the Brimer cluster analysis will lead to greater objectivity than the experimenter's interpretations of the childrens' performance, so that if there is a close correspondence between the results of the cluster analysis and the experimenter's interpretations (which will be in terms of the proposed models) this can be regarded as confirmation of the validity of the interpretations (within certain limits which will be elaborated later). The categorisation of subjects' responses used in the cluster analysis were also not subjected to any tests of inter-marker reliability, although evidence from other studies using the same technique (Satterly and Brimer, 1971; Wallace, 1972a) indicates that high consistency between markers may be expected.

In order to carry out the first part of the proposed analysis it is necessary to devise a scheme for classifying the results in terms of the various proposed models. The results are to be scored as type 1, 2, 3 or 4, corresponding to models 1, 2, 3 and 4 respectively. The criteria used in interpreting the results are as follows:

Type 1: Any answer corresponding to model 1, Piaget's first developmental stage. Such answers appear to the experimenter to involve spurious factors or guesswork. Choices are erratic and their justifications often seem post hoc. The following examples, taken from the experimental results, serve to illustrate this.

SB10 (Beads experiment subject number 10), problem B4 red
(beads experiment problem 4, target colour red.)
Chooses 6G4R (the box with six green and four red beads inside). 'It's got more sharper corners'.

SB7, problem B3 red.
Chooses 4G4R 'I like this box'.

SB4, problem B2 red.
Chooses 2G2R 'My best one. My Mummy told me'.

SB3, problem B2 red.
Chooses 2G3R 'To get the red out'.

SB13, problem B1 red.
Chooses 4G2R. 'There's two reds and two greens but they're on different sides'.

Table 1(a): Choices consistent with the use of type 2, type 3, and type 4 strategies in the first beads experiment.

	Collection A		Collection B	
Problem B1:	4G2R	v	2G4R	(G = green bead)
Type 2 choices:	Green target, 4G2R.		Red target, 2G4R.	(R=red bead)
Type 3/4 choices:	Green target, 4G2R.		Red target 2G4R.	
Problem B2:	4G3R	v	2G3R	
Type 2 choices:	Green target, 4G3R		Red target, either.	
Type 3/4 choices:	Green target, 4G3R		Red target, 2G3R.	
Problem B3:	4G4R	v	2G2R	
Type 2 choices:	Green target, 4G4R		Red target, 4G4R.	
Type 3/4 choices:	Green target, either		Red target, either.	
Problem B4:	6G4R	v	2G2R	
Type 2 choices	Green target, 6G4R		Red target 6G4R	
Type 3/4 choices	Green target, 6G4R		Red target 2G2R	
Problem B5:	8G4R	v	2G2R	
Type 2 choices	Green target, 8G4R		Red target, 8G4R.	
Type 3/4 choices:	Green target, 8G4R		Red target, 2G2R.	
Problem B6:	6G4R	v	4G2R	
Type 2 choices:	Green target, 6G4R		Red target, 6G4R.	
Type 3 choices:	Green target, either.		Red target, either.	
Type 4 choices:	Green target, 4G2R.		Red target, 6G4R.	
Problem B7	8G4R	v	4G2R	
Type 2 choices:	Green target, 8G4R		Red target, 8G4R.	
Type 3 choices:	Green target, 8G4R		Red target, 4G2R.	
Type 4 choices:	Green target, either.		Red target, either.	

SB16, problem B1 red.

Chooses 2G4R 'I got a green one out before'. (This was the outcome of her previous choice.)

Type 2: Any answer corresponding to model 2. Such answers involve a comparison of the numbers of target beads in each box, and choice of the box with more of them. Answers involving comparison of the numbers of non-target beads in each box (without reference to the target beads), and choice of the box with less, may also be scored as type 2, although these are quite rare. The choice made and reason given must not conflict, hence the choices will be as shown in Table 1 (a).

Examples of Type 2 responses are given below. (Ages of the subjects quoted are not given, because the examples are meant to be illustrative of the particular type, rather than any given age).

SB28, problem B4 green

chooses 6G4R 'Got six greens in. Other only has two.'

SB30, problem B4 green

chooses 6G4R 'More greens than the other box.'

SB87, problem B1 red.

chooses 2G4R 'There's two more reds than there' (points to other box.)

SB64, problem B1 green

chooses 4G2R 'Not so many red ones as the other box'. (Answers of this type are rare).

SB63, problem B1 red.

chooses 2G4R 'Twice as many reds as the other one.' (Only one subject did this, but it is worthy of comment in view of the fact that it involves comparison by ratios, with the type 2 strategy.)

In problem B2 there are an equal number of red beads in each box, so that when red is the target colour an additional distinction between responses generated by model 2(a) and responses generated by model 2(b) is necessary. Model 2(a) leads to assertions that there are no grounds for choosing between the two boxes, whilst 2(b) leads to choice of the box with less beads of the non-target colour, often preceded by a long silence or an assertion of type 2(a).

Examples of type 2(a) responses:

SB12, problem B2 red.

chooses 2G3R 'Don't know why. Both got three. Doesn't matter which you choose.'

SB41, problem B2 Red.

chooses 4G4R 'Just a guess. Same really, I think.'

(This can only be seen as 2(a) in the context of the subject's response to the other part of problem B2.)

Examples of type 2(b) responses:

SB44, problem B2 red.

chooses 2G3R ''Cos I've just said the other one. Three in each, but not the same. The greens, aren't the same. I'll choose the one with less greens.'

SB48, problem B2 red.

chooses 2G3R 'Same reds, but less greens.'

SB63, problem B2red.

chooses 2G3R 'Only two greens. Same reds in each one.'

Type 3: Answers corresponding to model 3. This involves comparison of the numbers of target and non-target beads in each box, and choice made and the reason given must be in agreement, so that the choices will be as shown in Table 1(a).

Examples of type 3 responses taken from the experimental results:

SB71, problem B3 red.

chooses 4G4R 'Doesn't matter - they've each got the same number of reds and greens.'

SB60, problem B7 green

chooses 8G4R 'There's four more, only two more in there.'

SB67, problem B4 red

chooses 2G2R 'Equal ones. More chance of green with the other.'

SB66, problem B2 green.

chooses 4G3R 'Got most greens and three reds. The other only has two greens and three reds.'

<u>Type 4:</u> This categorisation is given to answers, which appear to have been arrived at by the 'true proportion' model of Piaget's third developmental stage, one possible version of which is model 4. The choices made in problems B1, B2, B3, B4 and B5 will be identical with the choices made in accordance with model 3, but the reasons for the choices will be different. The choices in problems B6 and B7 will be different to those made by model 3 and are given in Table 1(a).

Examples of type 4 taken from the experimental results:

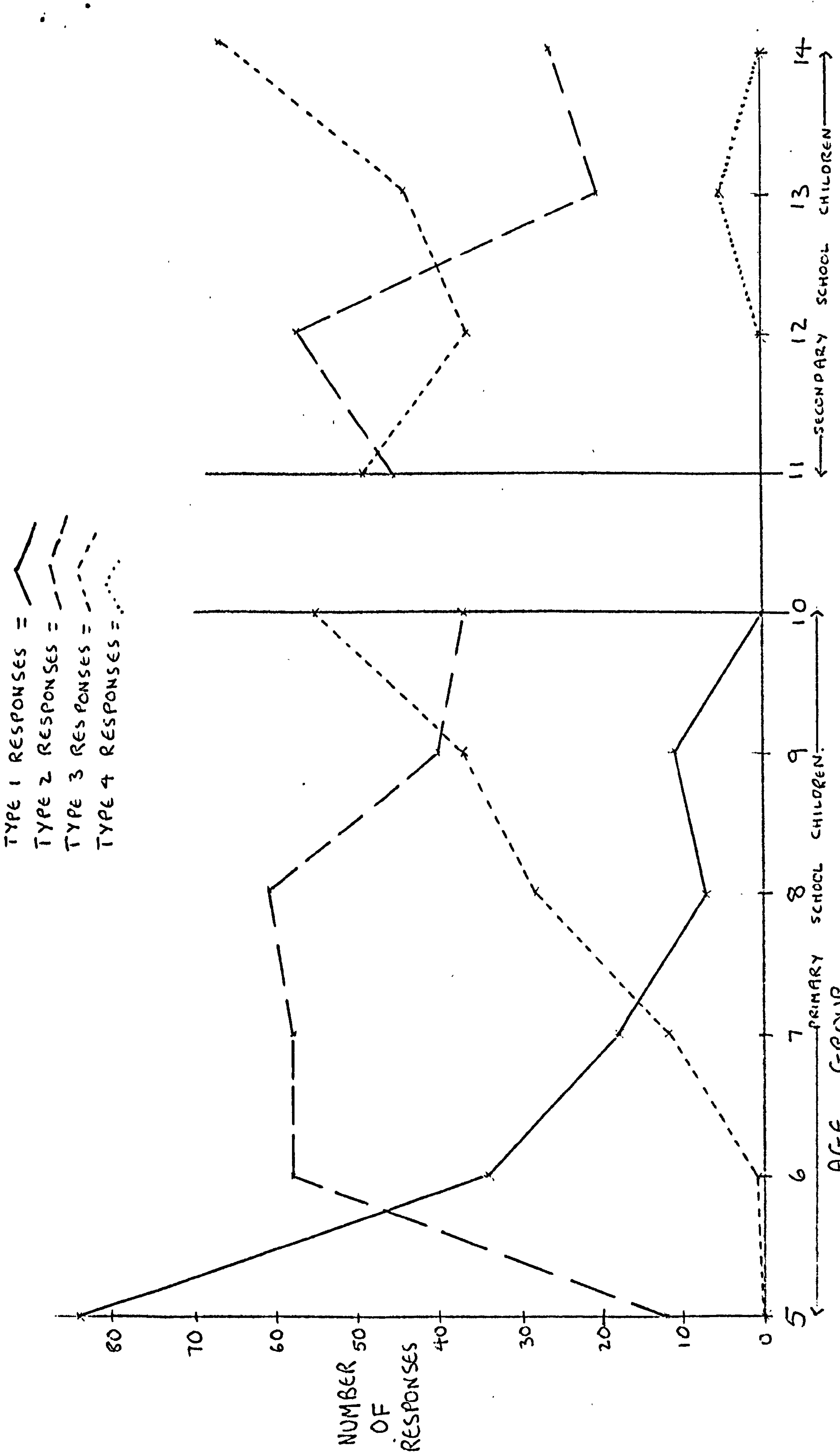
SB 108, problem B1 green.

chooses 4G2R 'There's four greens and two reds; a two to one chance.'

SB107 , problem B6 green.

chooses 4G2R 'There's more greens and reds in the other, less here. Two to one change here, one and a half to one there.'

FIGURE 4: GRAPHICAL ILLUSTRATION OF THE RESULT OF THE FIRST BEADS EXPERIMENT IN TERMS OF THE SOLUTION STRATEGIES USED AT DIFFERENT AGES.



SB108, problem B4 red.

chooses 2G2R. 'It's got the same chance as green. Better chance than in the other one. Halves each.'

In addition to these types of responses, a certain number were found to be unclassifiable within the scheme, and these are grouped together as type '?' These include answers which supply too little information to be classed as type, 2, 3 or 4, but too much to be called type 1, and answers where the model 2 or model 3 strategy is used but the box with less instead of more of whatever is required is chosen.

Examples of unclassifiable responses:

SB51, problem B4 red.

chooses 2G2R, 'Only two red ones there, but four in the other.'

Experimenter: 'What do you mean by that?'

SB51: 'Don't know. I'm sure its better.'

SB35, problem B3 red.

chooses 2G2R. 'Got less. Got two red and two green.'

The results of applying this categorisation scheme to the data can be found in Appendix B, and a graphical illustration is provided in Figure 4. In order to categorise the results each problem was treated separately, but information was exchanged between problem subparts to resolve some ambiguities. For example, if a child gave an answer to one part of problem B1 which was clearly of type 2, and an answer to the other part which could be seen as either type 2 or type 3, this was considered to be of type 2. Information was only exchanged in this way between subparts of the same problems, i.e., within problems. If this is not done the number of ambiguities in the results rises. In order to avoid any temptation to treat the subparts of separate problems in this way the results were scored by scoring all the answers to a particular problem at a time, rather than scoring all the answers a particular child gave to different problems at one time. In this way 'halo' effects are eliminated in all but the within-problem case. In the Brimer analysis every subpart of every problem is treated separately.

<u>Discussion</u>: The ease with which the classification scheme derived from the theoretical models can be applied to the data, and the small number of unclassifiable responses, provides subjectively convincing evidence for the validity of the models as representations of the way children are handling the problems. The graph (figure 4) shows that models 1, 2 and 3 represent typical ways of dealing with problems

of this type, but model 4 does not, (all the type 4 responses to problems B1 - B4 were provided by one subject), and may be more closely related to the subjects' mathematical sophistication.

The results from the secondary shool children show something of a fall-off in comparison with the primary school results. There are various reasons why this might have been expected. Although the catchment area of the primary and secondary schools involved is the same, some children will have been 'creamed off' and sent to other schools in the area. The attitude of the secondary school children to the experiment is one of polite condescension as compared with the enthusiastic involvement of the primary school children, and they gave the impression of being unwilling to take it seriously. Because of this the analysis will concentrate on the results of the primary school group and any conclusions drawn from the secondary school results can only be tentative.

The overall result does not accord with Piaget and Inhelder (1951) who do not postulate anything corresponding to model 3, and consider model 4 as the final stage of conceptual development in this area. However, their experimental design would not have allowed models 3 and 4 to be distinguished, so that they may well have mistaken answers generated by model 3 for answers involving a true appreciation of proportion. Their conclusion that model 4, which corresponds to their third developmental stage, is the natural outcome of formal operational reasoning, may not be invalidated by the results of the present study because of the reasons cited for believing that the performance of the secondary school children involved was not all it might have been.

The research reported by Lowe and Ranyard (1973) also fails to distinguish models 3 and 4. In fact their idea of a proportion strategy seems to bear some similarity to model 3, as is shown by some of the statements already singled out from their report.

The graph of the overall result (figure 4) gives preliminary evidence of a developmental sequence from model 1 to model 2 to model 3 (if allowance is made for the drop in performance between the primary and secondary school children) with the different types of solution showing maxima at different ages. This raises the possibility that the models are in fact models of 'developmental stages' in the solution of this type of problem. Before this can be decided the nature and meaning of the term stage must be clarified.

Considerations concerning the detection of stages in cross-sectional data.

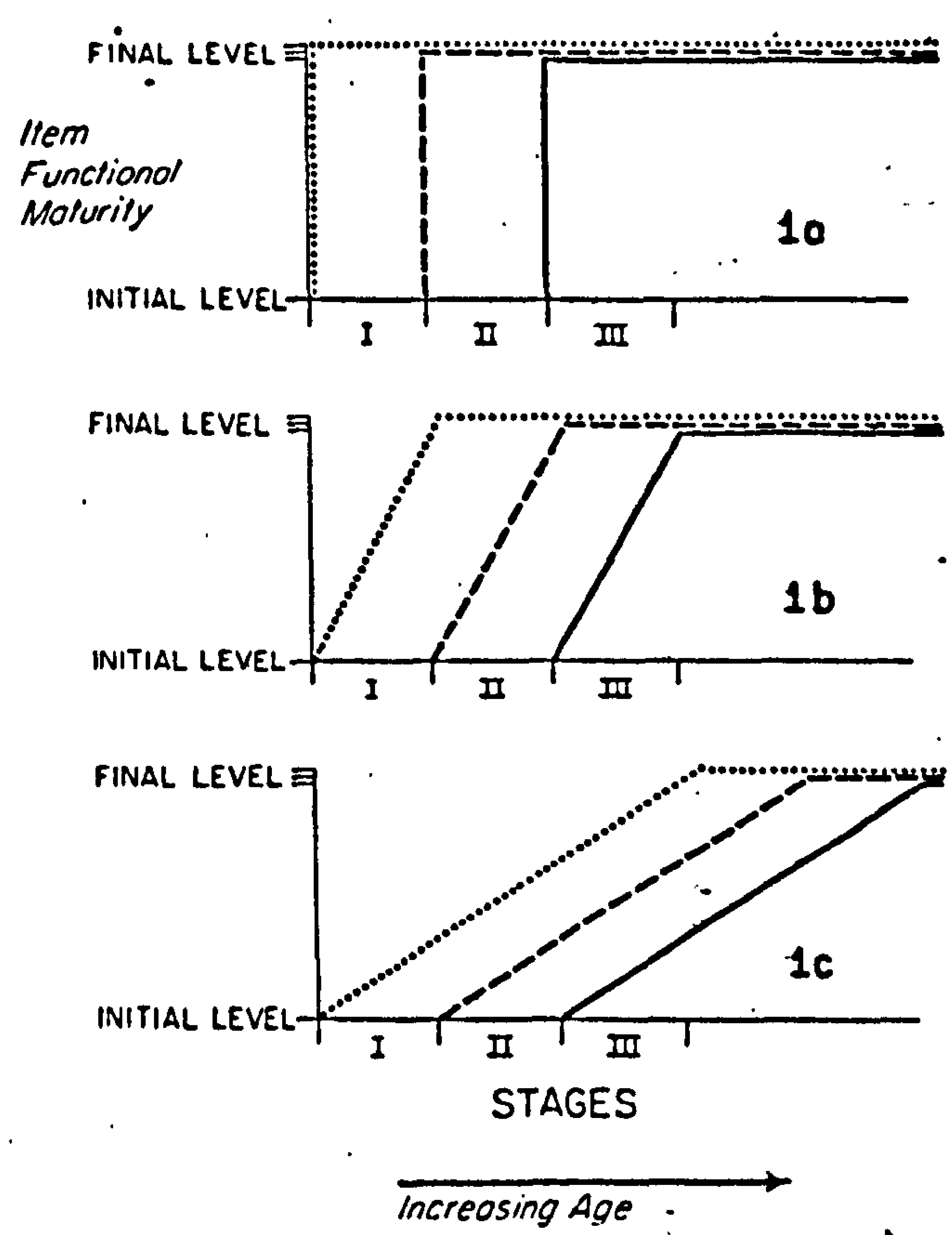
Stage has had a long history as a psychological concept. Much of this history is one of spurious usuage for, as Wallace (1972a) remarks, an expression such as 'He's at the teething stage' can be reduced to 'He's teething' without any loss of meaning. The most interesting use of the term has been made by Piaget, who sees stages in terms of qualitative differences in intellectual organisation and, consequently, behaviour. In addition to the stress on qualitative changes, each stage is seen in a wider framework as part of a sequence of increasingly structured and integrated adaptations to the environment. Authoritative expositions of Piaget's views are provided by Piaget (1960) and Inhelder (1962),

> 'Whereas somatic and perceptual development seem to be continuous, intellectual development seems to take place in stages, the criteria of which can be defined as follows:
>
> 1. Each stage involves a period of formation (genesis) and a period of attainment.. Attainment is characterized by the progressive organisation of a composite structure of mental operations.
> 2. Each structure constitutes at the same time the attainment of one stage and the starting point of the next stage, of a new evolutionary process.
> 3. The order of succession of the stages in constant. Ages of attainment can vary within certain limits as a function of factors of motivation, excercise, cultural milieu, and so forth.
> 4. The transition from an earlier stage to a later stage follows a law of implication analogous to the process of integration, preceding structures becoming a part of later structures.' (Inhelder, 1962, p. 24.)

This approach to stages in conceptual development has been elaborated by Pinard and Laurendeau (1969), who also provide a discussion of the value of Piaget's equilibration model as a model of the factors responsible for the transitions between stages. The problem of the nature of such a transition mechanism is of fundamental importance in developmental psychology and will be approached later. At the moment it is necessary to establish some sort of practical prescription for the detection of stages in a set of data.

The theoretical side of this problem has been considered very thoroughly by Flavell (1971). Flavell takes the various criteria of

FIGURE 5: THREE MODELS OF THE DEVELOPMENTAL COURSE OF INDIVIDUAL STAGE-SPECIFIC COGNITIVE ITEMS (FIGURE 1 FROM FLAVELL, 1971, P.426).



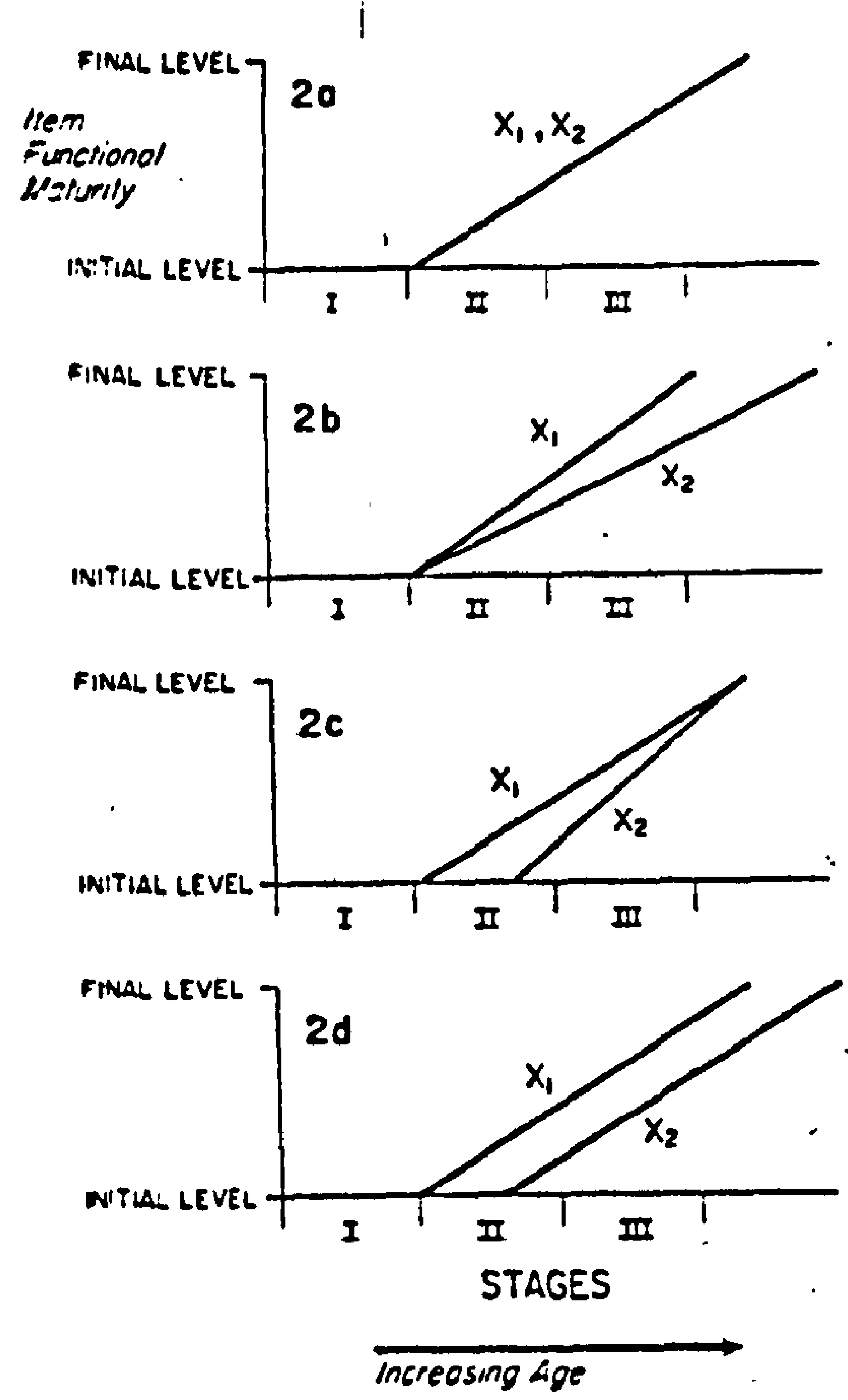
a stage development in turn and analyses the way in which these might influence performance in a way which would allow them to show up in experimental data. His starting point is that any meaningful use of the term stage must involve qualitative rather than quantitative changes in thinking. One would not, for example, want to say something like 'This child had previously been in the stage where his short-term recall was poor, but is now in the stage where it is better.' (Op.cit.,p.425.) Unfortunately, this does not mean that qualitative and quantitative changes cannot co-exist in the developmental process. There is abundant evidence of quantitative-looking improvements in the development of perception and memory (see Gibson, 1969, Flavell, 1970b)and Pascual-Leone (Pascual-Leone and Smith, 1969; Pascual-Leone, 1970) has claimed that such quantitative developments may account for the transition from one Piagetian stage to the next. It is the apparent mixture of quantitative and qualitative changes in conceptual development, of improved skills and novel strategies, which makes the firm detection of stages so difficult.

Suppose, for example, we have an 'ideal' stage development sequence in which each stage can be detected by the presence or absence of a single item. Possible ways in which this sequence might develop are indicated in Flavell's figure 1 (Flavell, 1971, p. 426), which is presented here as figure 5. (The abscissae in figure 5 show the defined age periods for three developmental stages. The linear development is an arbitrary convention.)

A purely qualitative change model would be represented by 1a in figure 5. This implies that there are no periods of transition between stages, (or, moreprecisely, that the transition periods are of null duration), leading to the paradoxical conclusion that the child spends all his time 'being' rather than 'becoming'. To most people such a view would be unacceptable, and some notion corresponding to 'functional maturity', which is a measure of how readily available the item is in the child's repertoire, would have to be introduced. If the full functional maturity of a given item is taken to mark the end of one stage and the beginning of the next, model 1b is produced. It is this model which appears to be implicit in the efforts of many Piagetian researchers. However, a more extreme version of the gradual development model, 1c, is possible in which the final level of functional maturity of a stage-specific item is not reached until after development of the next stage item has begun. Flavell maintains that for most items so far investigated this model gives the most realistic picture.

45A.

FIGURE 6: FOUR POSSIBLE INTERPRETATIONS OF 'DEVELOPMENTAL CONCURRENCE' BETWEEN TWO SAME-STAGE COGNITIVE ITEMS. (FIGURE 2 FROM FLAVELL, 1971, p. 437).



When attention is shifted from ideal stages defined by single items to more realistic stages (from the Genevan point of view at any rate) involving various items, the situation becomes even more confused and confusing. The various items defining a stage may show varying degrees of concurrency in their development. Taking the case of a single stage defined by two items, X1 and X2, Flavell draws attention to the possibilities illustrated in his figure 2, (Op.cit.,p.437), presented here as figure 6.

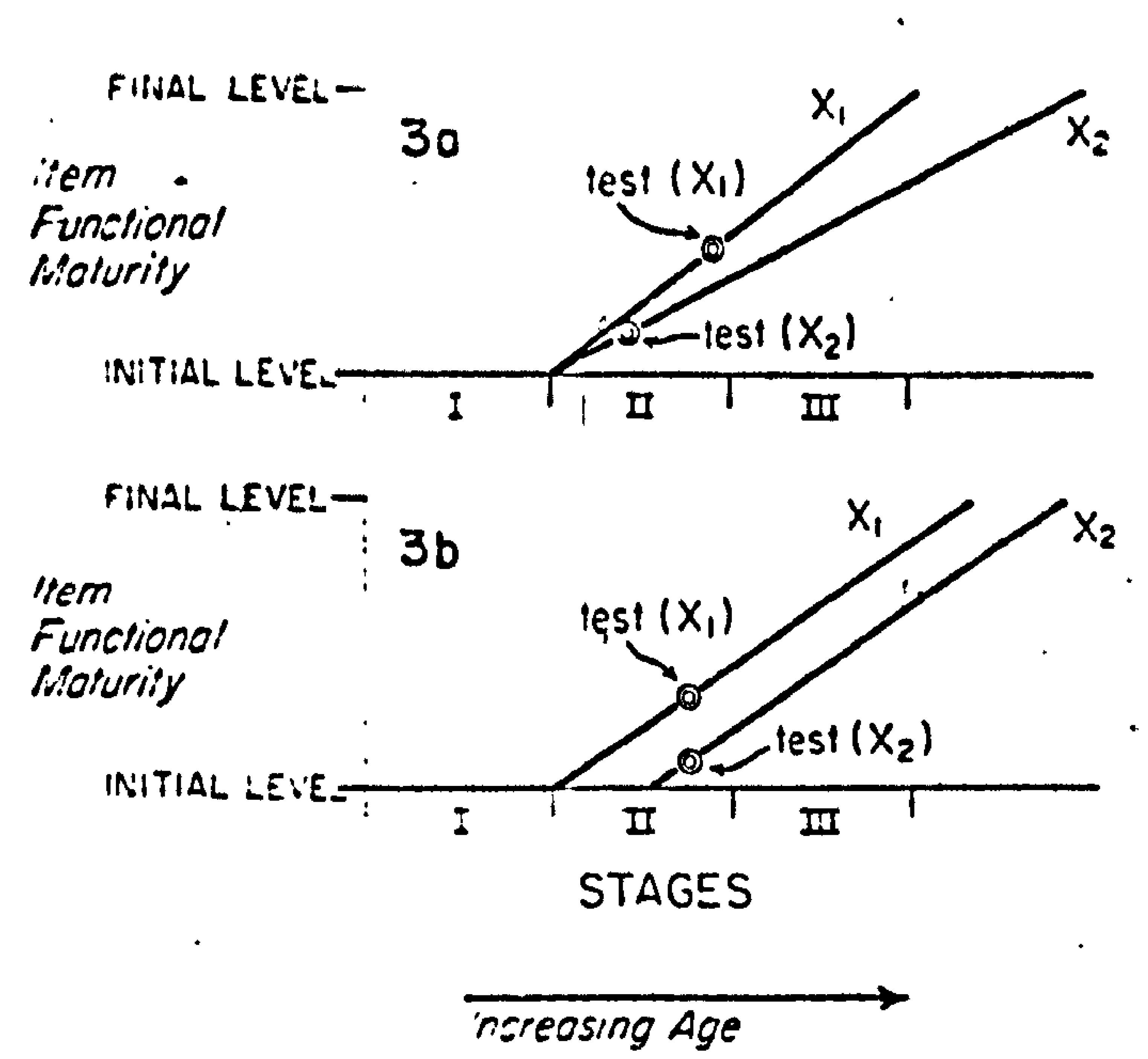
Once again the linear development is purely conventional. The important thing, it is argued, is whether the items X1 and X2 are initiated or terminated concurrently. In this case 2a would seem to be most like the Genevan conception of stage development, although 2b and 2c are not wholly incongruent with it.

As soon as the possibilities outlined in figure 5 are combined with those of figure 6 the situation becomes very confusing indeed. The result is brilliantly summarised by Flavell,

> 'It is clear that "developmental concurrence" can be an ambiguous expression once the acquisition of an item is regarded as an extended process (figures 1b and 1c) rather than an instantaneous, quantum-like jump (figure 1a). In particular, to say that "there is (exists) developmental concurrence" between two acquisition processes does not differentiate among Models 2a - 2c. Nor does it distinguish these from 2d, nor, in fact, 2a - 2d from the different-stage developments represented by the dotted, solid and dashed lines in Figure 1c (the stage I item is after all, still developing during a time interval in which the stage III item is developing, and hence the two are for a time "developing concurrently").
>
> What specific type or degree of concurrence do people have in mind, then, when they talk about items of the same stage developing "simultaneously" or "together"? The answer is, not surprisingly, that they generally do not have anything very specific in mind, because the distinctions represented in Figures 1 and 2 have not entered into their thinking on that matter.' (Op.cit.,p.438).

The difficulty of deriving evidence concerning what is actually happening in development in order to resolve the kind of issue raised above is highlighted by Flavell. The usual method of testing for the presence of developmental concurrences has been to administer tests designed to detect the presence or absence of items X1 and X2 to a group of children(Flavell, 1970a, Wohlwill, 1963). The

FIGURE 7: SOURCES OF ERROR IN THE DIAGNOSIS OF DEVELOPMENTAL CONCURRENCE - NONCONCURRENCE OF TWO SAME-STAGE COGNITIVE ITEMS.
(FIGURE 3 FROM FLAVELL, 1971, p.440).



pattern of results obtained can then be interpreted as indicating the concurrence or non-concurrence of X1 and X2. The difficulty with this method is that if the criterion of success on each test (in terms of the level of functional maturity of the item required to pass) is not identical almost any interpretation can be made. Figure 7 (figure 3 in Flavell, 1971) shows two examples of this, in 3a the result is an apparent developmental asynchronism (a falsely negative finding), in 3b it is the converse (a falsely positive finding). (Flavell, 1971, p. 440).

The obvious solution to this particular problem is to equate the level of difficulty of the tests employed, but it is not clear how far this could be realised in practice. Flavell's own solution, which is also a solution to the more general problems he has posed, is to 'liberalise' the concept of stage and argue, contrary to Pinard and Laurendeau (1969), that it can be used meaningfully without really requiring concurrences. His conclusions are:

> 'The items that define a stage develop gradually rather than abruptly. Moreover, the typical item probably does not achieve its "final level of functional maturity" (defined in terms of the item's evocability and utilizability as a solution procedure) until after the conventional termination age of the stage in which it is supposed to begin its development.' (Flavell, 1971, p 450.)

With respect to inter-item concurrence:

> 'First, items from the same stage may often emerge in an invariant or near-invariant sequence rather than concurrently, although important methodological problems cloud the research evidence on this point. Second, a stage theory such as Piaget's does not in any event logically require anything but a very loose sort of item concurrence at most, and research attempts at establishing strict concurrences have been accordingly misguided in rationale'. (Op.cit., p.451.)

Flavell's discussion of the stage concept has been accorded an extended treatment because it seems well-balanced between theoretical and practical perpectives. There are two points which should perhaps be added. The first is that Piaget's own use of the term is itself inconsistent. He is very definite about how it is to be defined (e.g. Piaget, 1960), but a careful reading of his own definition shows that only the 'major' developmental stages (the sensorimotor, concrete operational and formal operational periods) conform to it. The other stages Piaget identifies depart from the ideal in varying degrees. In Piaget (1951), six stages in the development of imitation are outlined. In Piaget (1952) there are three stages in the development of conservation of discontinuous quantities, three stages in the development of conservation of continuous quantities, and three stages in the development of additive composition of classes. In Piaget, Inhelder and Szeminska (1960) are three stages in the child's understanding of 'changes in position', three stages in the development of spontaneous measurement, three stages in the construction of relations of distance, three stages in the development of conservation of length, three stages in the development of measurement of length, three stages in subdividing a straight line, three stages in locating a point, three stages in angular measurement, three stages in measuring the angles of a triangle, four stages in summing the angles of a triangle, three stages in the development of solutions to problems of geometrical loci, four stages in the representation of curves of movement, three stages in the conservation of area, three stages in the measurement of area, three stages in the subdivision of areas, four stages in doubling an area or volume, and four stages in the conservation of volume, to say nothing of 'substages'.

Frequently the three-stage developments are little more than statements that there is initial absence of something, followed by a period of transition, followed by its attainment. In other cases, such as the three-stage development from Piaget and Inhelder (1951) which has been quoted many times in this thesis, the stages are just different answers children give to the same question. This is not to deny that they are 'genuine' stages, or to invalidate Piaget's claims, but such stages can hardly pass the complex set of criteria put forward by Inhelder (1962). The gap between this cavalier empiricism and the highly systematised Piagetian theory is most striking, and supports Flavell's contention that not all of Piaget's theoretical definition is essential to his concept of stage. Of course Piaget's own concept of stage has 'developed' in the years, and recent discussions (Inhelder,1972; Cellérier,1972) indicate that it may well be moving in the same direction.

The second point is that it seems to the writer that much of the confusion in this field can be avoided if a simple but radical step is taken. This is to admit that the inferences concerning stages which can be derived from experimental data and the inferences to be derived from theoretical models are of a different order. Inferences from experimental data can only be of an 'as if' variety, the use of the term stage in the description of empirical results is an analytic convenience. The problem of distinguishing qualitative and quantitative phenomena solely on the basis of empirical data is intractable, as the sophisticated consideration Flavell has given it reveals. On the other hand any model of children's performance which passes the criteria of sufficiency (Gregg and Simon, 1967) and developmental tractability (Klahr and Wallace, 1970) will allow definte conclusions regarding stages and the role of quantitative and qualitative factors to be drawn.

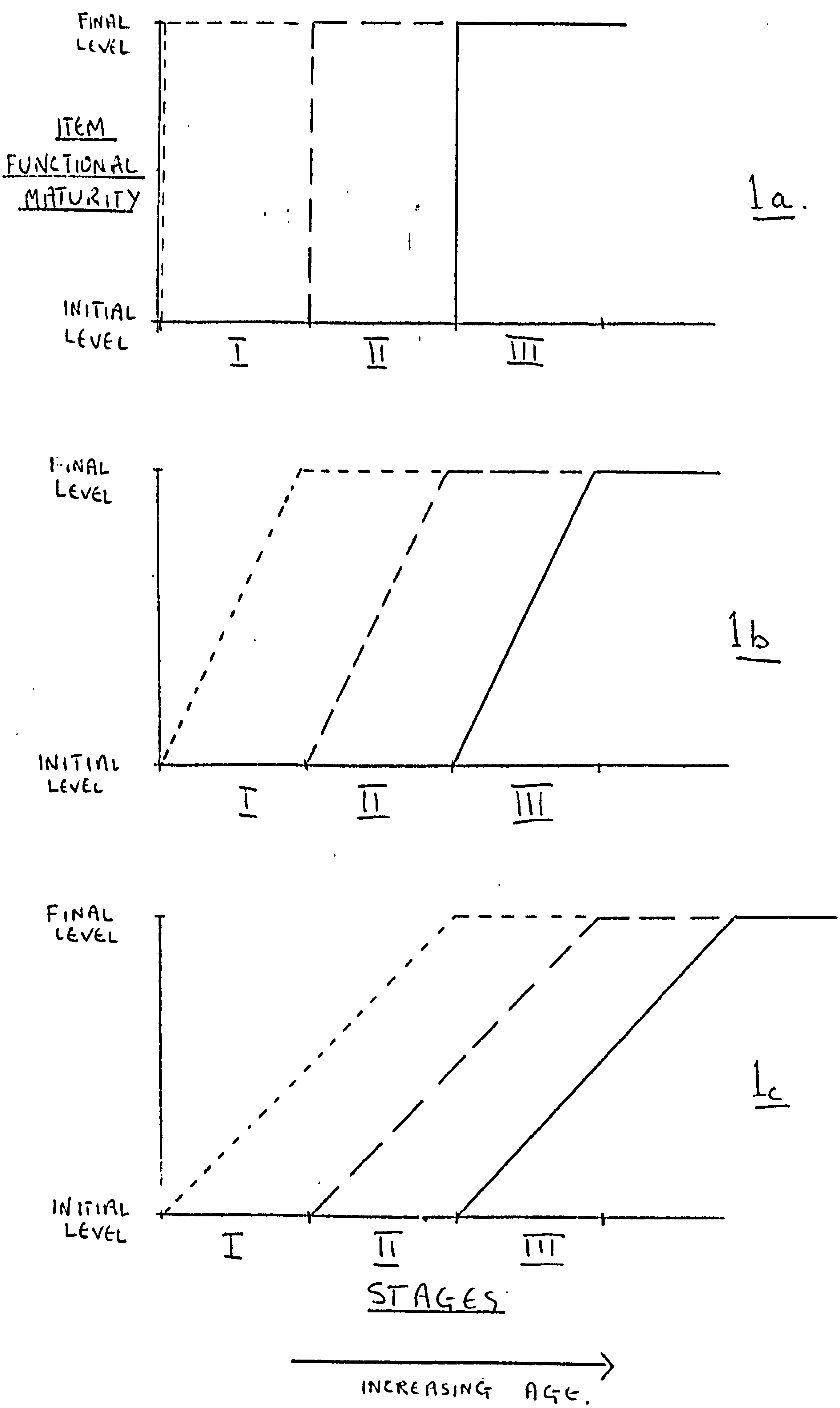
Acceptance of this position does not mean that 'stage' is to be abandoned as a theoretical concept, but that the support for its use must come from theoretical and not just empirical considerations. When adequate theoretical models are available it is possible to compare experimental results generated in different situations with predictions derived from the models. There cannot be a purely objective way of deriving stages from experimental results, but certain patterns of experimental data are consistent with certain stage models, whilst other patterns are not. Of course, our models need not involve stages at all, but those being discussed here are of the stage type, so that the kind of data which would be consistent with the stage model under consideration must be elucidated.

The models of problem solving strategies presented in this thesis can be viewed as a sequence of developmental stages in the looser sense used by Piaget, of developmental stages in the solution of a particular type of problem. They do not meet the criteria of stages proposed by Piaget (1960) or Inhelder (1962), which define more general stages in cognitive development. Now obviously stages in problem solving strategies are not unrelated to more general stages of cognitive development, but they are different in certain crucial ways. The main difference is with respect to what Flavell calls inter-item concurrence. Any developmental stage of the order of complexity described by Piaget (1960), or Inhelder (1962), will have to be detected by the presence or absence of several items, and the different ways in which these stage-related items might develop presents a major difficulty in stage detection. However, developmental stages in problem-solving strategies will often only involve the presence or absence of a single item, namely the strategy under consideration, and this greatly simplifies the problem of comparing theoretical predictions and results. If this arguement is valid we arrive at the paradoxical position that one type of developmental stage is defined by several features but is difficult to detect with any degree of certainty, whereas the other type is defined by less features and yet can be comparatively easily detected.

The problem of defining stages is thus sidestepped by the claim that stages are the result of a particular way of modelling the developmental process, and that if the proposed models meet the criteria of sufficiency and developmental tractability then the type of stages they will lead to can be easily deduced. The models 1, 2, 3 and 4, of possible strategies in the solution of probability-comparison problems meet these criteria and would predict a stage-development in childrens' solutions of these problems. This stage-development will be in the direction 1 - 2 - 3 - 4 but whether each stage must be passed through, or some may be by-passed, is not clear without a model of the transition mechanism responsible for the development. For the present purposes this is unimportant as the data to be considered is cross-sectional rather than longitudinal. In a longitudinal study it would be most important to decide whether the stage development is to be considered as an invariant or only semi-invariant sequence, but a cross-sectional study is by its nature insensitive in this respect. The presence or absence of a given stage can be detected in each problem by a single item, namely the presence or absence of a verbal report of the solution-strategy characteristic of that stage. This means that the stage-development, if it is found, will be like one of the possibilities illustrated in figure 5 (Flavell's figure 1).

A practical means must now be found for distinguishing these possibilities empirically, and finding out which one the data of the first beads experiment supports. One solution to this would appear to be to use the concept of 'stage mixture' outlined by Turiel (1969). In a discussion of the developmental processes which might be involved in childrens' moral thinking, Turiel points out that Piaget's view of stages as 'structural wholes can give the impression that, at any given time, a child functions on one stage and that change involves movement from one such discrete stage to another. (This is model 1a in figure 5) In practice, however, Turiel found that normally a child's profile of scores in different situations contains a dominant stage having the largest scores, with the more distant stages from the dominant stage having smaller scores. Turiel proposes that this kind of stage mixture is an essential part of development and may be directly related to the structuring process.

FIGURE 8: A MODIFIED VERSION OF FLAVELL'S (1971) FIGURE 1, TO ILLUSTRATE AN 'INTEGRATIVE' STAGE DEVELOPMENT.



That is, periods of change might be characterised by extensive stage variation, while periods of fixity would be characterised by consistency.

These ideas are, of course, very similar to those illustrated by Flavell in the figure given here as figure 5. Turiel's contribution is to point to the possible use of stage mixture as a criterion for distinguishing the different stage development models which were later clarified by Flavell (1971). Consideration of Flavell's figure and the patterns of stage-mixture which would be exhibited by each of its three models shows that they can indeed be separated by this method. To illustrate this figure 5 must be considered again. Notice that when any item attains its full functional maturity its development line continues parallel to the horizontal axis. However, one of the main features of Piaget's idea of stage development is that preceding stages are integrated into subsequent stages. This means that when item II attains its full functional maturity, item III will be incorporated into it. With the developmental stages for the strategies proposed here a similar effect will occur, since when any stage is fully developed there will be no call to use the solution strategy of the preceding stage and this will effectively disappear. Hence a more appropriate version of the figure would be that shown in figure 8.

It can of course be pointed out that such an 'integrative' notion of stage-development need not be used, (see e.g.Werner, 1948,who maintains that each stage passed through is retained and accessible, although lower stages are subordinated to the higher stages, or Flavell, 1972, where further distinctions for possible developments of stage-related items are drawn). However, the models of problem-solving strategies being considered here are of this type, and it has been argued that the notion of stage is only useful in conjunction with some sort of explicit model. In cases where each stage becomes integrated into the subsequent stage in this way it can be seen that the degree of stage-mixture obtained empirically does indeed distinguish the different possible kinds of development. For example:

Flavell's model 1a involves no stage mixture.

Flavell's model 1b involves stage mixture only between one stage and the stage either preceding or following it. In other words, a child is in a stage and can either regress or improve on certain problems, but no child will show both regression and improvement on different problems. Put more clearly, no child will exhibit more than two different stages in the whole range of problems used.

Flavell's model 1c involves stage mixture between one stage and the stages both preceding and following it. A child may exhibit

51B

TABLE 2 (CONT'D).

SUBJECT	AGE	PROBLEM B1		PROBLEM B2		PROBLEM B3		PROBLEM B4		
		GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	
SB41	8	2	2	2	2(a)	2	2	2	2	
SB42	8	2	2	2	2(a)	2	2	2	2	
SB43	8	2	2	2	2(a)	2	2	2	2	
SB52	9	2	2	2	2(a)	2	2	2	2	(STAGE 2).
SB61	10	2	2	2	2(a)	2	2	2	2	
SB62	10	2	2	2	2(a)	2	2	2	2	
SB32	7	2	2	2	2(b)	2	2	2	2	
SB23	6	2	2	2	2(b)	2	2	2	2	
SB33	7	2	2	2	2(b)	2	2	2	2	
SB53	9	2	2	2	2(b)	2	2	2	2	
SB63	10	2	2	2	2(b)	2	2	2	2	
SB64	10	2	2	?	?	2	?	2	2	
SB24	6	2	2	2	3	2	2	2	2	
SB44	8	2	2	2	2(b)	3	3	2	2	
SB34	7	3	3	2	3	2	?	?	?	
SB54	9	3	2	2	2(b)	3	3	2	?	
SB55	9	3	3	?	?	2	?	3	3	
SB71	10	2	2	3	3	3	3	2	2	
SB36	7	3	3	3	3	2	2	2	2	
SB48	8	3	3	2	2(b)	3	3	2	3	TRANSITION, 2-3.
SB35	7	3	3	3	3	2	?	3	?	
SB56	9	3	3	3	3	2	2	3	2	
SB65	10	3	3	3	3	2	2	3	2	
SB66	10	3	3	3	3	?	3	3	2	
SB57	9	3	3	3	3	2	2	3	3	
SB58	9	3	3	3	3	2	2	3	3	
SB45	8	2	2	3	3	3	3	3	3	
SB59	9	3	3	3	3	2	2	3	3	
SB46	8	3	3	3	3	3	2	3	3	
SB60	9	3	3	3	3	3	2	3	3	
SB67	10	3	3	3	3	3	3	3	3	
SB68	10	3	3	3	3	3	3	3	3	
SB47	8	3	3	3	3	3	3	3	3	
SB69	10	3	3	3	3	3	3	3	3	
SB70	10	3	3	3	3	3	3	3	3	STAGE 3.
SB72	10	3	3	3	3	3	3	3	3	

51A

TABLE 2: RESULTS OF THE FIRST BEADS EXPERIMENT ARRANGED TO SHOW THE STAGE TREND.

SUBJECT	AGE (YRS.)	PROBLEM B1 GREEN	PROBLEM B1 RED	PROBLEM B2 GREEN	PROBLEM B2 RED	PROBLEM B3 GREEN	PROBLEM B3 RED	PROBLEM B4 GREEN	PROBLEM B4 RED	
SB1	5	1	1	1	1	1	1	1	1	
SB2	5	1	1	1	1	1	1	1	1	
SB3	5	1	1	1	1	1	1	1	1	
SB4	5	1	1	1	1	1	1	1	1	
SB5	5	1	1	1	1	1	1	1	1	
SB6	5	1	1	1	1	1	1	1	1	STAGE 1.
SB7	5	1	1	1	1	1	1	1	1	
SB8	5	1	1	1	1	1	1	1	1	
SB9	5	1	1	1	1	1	1	1	1	
SB10	5	1	1	1	1	1	1	1	1	
SB13	6	1	1	1	1	1	1	1	1	
SB25	7	1	1	1	1	1	1	1	1	
SB49	9	1	1	1	1	1	1	1	1	
SB26	7	1	1	1	1	1	1	1	1	
SB37	8	1	1	1	1	1	1	2	1	
SB14	6	1	1	2	1	2	1	1	1	
SB15	6	2	2	1	1	1	1	1	1	
SB16	6	1	1	1	1	2	1	2	1	TRANSITION,
SB17	6	2	2	1	?	1	1	1	2	1-2.
SB27	7	1	1	2	?	?	?	2	2	
SB11	5	1	1	1	1	2	2	2	2	
SB51	9	1	1	2	2(a)	2	?	2	?	
SB18	6	2	2	1	1	2	2	2	?	
SB50	9	2	2	2	?	2	?	2	1	
SB19	6	2	2	2	1	2	2	2	?	
SB20	6	2	2	2	2(a)	2	?	2	?	
SB12	5	2	2	2	2(a)	2	2	2	2	
SB21	6	2	2	2	2(a)	2	2	2	2	
SB22	6	2	2	2	2(a)	2	2	2	2	
SB28	7	2	2	2	2(a)	2	2	2	2	
SB29	7	2	2	2	2(a)	2	2	2	2	
SB30	7	2	2	2	2(a)	2	2	2	2	
SB31	7	2	2	2	2(a)	2	2	2	2	
SB33	8	2	2	2	2(a)	2	2	2	2	STAGE 2.
SB39	8	2	2	2	2(a)	2	2	2	2	
SB40	8	2	2	2	2(a)	2	2	2	2	

(CONTINUED OVERLEAF).

three (or more if there are more than three stages) different stages in a range of problems.

This, at last, provides a workable way of interpreting the empirical data which has been classified in terms of the four models of strategies for solving the experimental problems presented earlier. In addition to the degree of stage mixture, there is one other criterion of a stage-development, which is that there should be a broad age-stage relation in the results, so that the older children belong in the more 'advanced' stages.

Developmental stages in the results of the first beads experiment.

Returning to the results of the first beads experiment, it can be seen that when each result is categorised as an example of one of the four problem-solving strategies derived from the theoretical models, the age-stage relation shows up very clearly. Figure 4 shows that type 1 responses are predominant at age five, type 2 are maximal between ages 6 and 8, and type 3 are predominant after age ten or eleven. Type 4 responses do not appear until age 12, and then are very infrequent.

In order to pursue this stage-development possibility further a table of the results has been drawn up in which the subjects are arranged in what would be increasing order of development if the stage-hypothesis were correct. This is table 2. Only subjects from the primary school group have been included in this table because there are reasons for supposing that the primary and secondary school group are not equivalent in certain respects. These reasons have already been discussed, and require that the two groups be kept apart as much as possible in the analysis. Inspection of the secondary school results (in Appendix B) shows that a similar table to table 2 could be made from them if required.

The groupings in table 2 are derived from the primary school childrens' responses to the experimental problems B1 - B4. Results of the additional problems, B5, B6 and B7 are not included in the table. The outcome of this grouping of the results is a three-stage development, with two transitions, which can be summarised as follows:

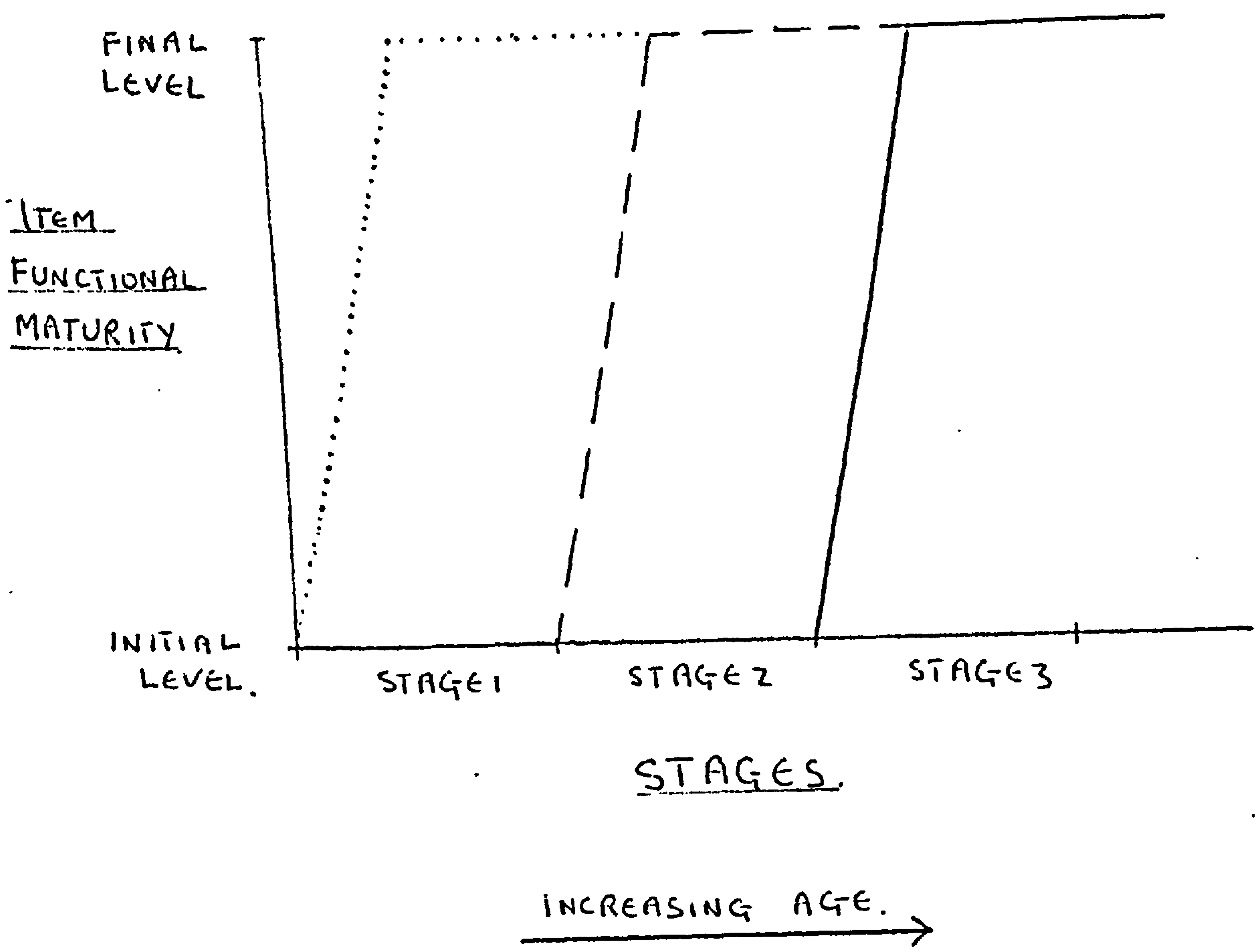
Stage 1: Fourteen subjects who gave only type 1 responses, mainly aged 5 years.

Transition 1-2: Twelve subjects who gave a mixture of type 1, type 2, and unclassifiable (?) responses, mainly aged 6 years.

Stage 2: Twenty-one subjects who gave only type 2 responses, mainly aged 7 and 8 years.

Transition 2-3: Nineteen subjects who gave a mixture of type 2, type 3 and unclassifiable responses. Mainly age 8, 9 and 10 years.

FIGURE 9: A MODEL OF A STAGE DEVELOPMENT CONSISTENT WITH THE RESULT OF THE FIRST BEADS EXPERIMENT.



Stage 3: Six subjects who gave only type 3 responses, mainly aged 10 years.

There were no subjects who gave mixtures of type 1 and type 3 responses, or mixtures of type 1, type 2 and type 3. All the unclassifiable responses came from children in the transition periods.

Following Turiel's (1969) discussion of stage mixture, these results are consistent with a three-stage development. The development corresponds to something in between Flavell's (1971) models 1(a) and 1(b), in which stage-mixture occurs only in the transitions between stages, and no more than two stages can be'mixed'. It is illustrated in figure 9.

It must be emphasised that this stage delineation can only be supported within the limits imposed by the methods employed. Flavell, for example, stresses that functional maturity can only be inferred from a wide range of situations, and it is quite possible that the range represented by problems B1 - B4 may not be wide enough. This would mean that if the range of difficulty encompassed by the problems were to be increased, some subjects might move from their present position within stages 2 or 3 into the transitions. However, major shifts, such as from stage to stage, should not occur.

This kind of effect can be seen in the answers to problem B6 and B7. Problem B6 and B7 were designed to investigate whether the model 3 strategy is used by children in a carefully controlled situation, and in general the results show that it is. However, in most cases children who have given type 3 answers to all parts of problems B1 - B4 show 'regression' to type 2 answers in some part or parts of problems B6 and B7 (see Appendix B). Of the six subjects in the primary sbhool group who were placed in stage 3, five showed this kind of regression.

What is causing this? The answer seems lie in Flavell's concept of functional maturity. Both problem B6 and problem B7 present an increased level of difficulty over problems B1 - B4, because of the larger numbers of beads involved. However, problem B7 involves the most beads and only one out of the six stage 3 subjects regressed when faced with this problem, whereas five out of six regressed whên faced with problem B6. Problem B6 is the problem on which model 3 leads to the prediction that either box can be chosen because there are two more greens on each side. If the model 3 type of reasoning is not fully functionally mature then a tendency to go along with the experimental set and make a choice is to be expected, and the next best way of making a choice becomes the model 2 type of reasoning.

This explanation accounts for the regression quite neatly and can be pursued further. Problem B3 is similar to problem B6 in that model 3 leads to the choice of either box, and inspection of the results shows that in the transition between stage 2 and stage 3, a stage 2 response is more likely to be given to this problem than to problems B1, B2 or B4. Of course the argument that difficulty over choice leads to regression could equally be turned on its head, i.e. difficulty over choice might lead to improvement. Within the primary school group this was not found to be the case, but of course the older children in this group are only just moving into stage 3, and some evidence of improvements on difficult problems is noticeable in the secondary school group.

Perhaps improvements might be more profitable sought in the cases where the model 2 type of reasoning runs into difficulties, such as problem B2. The answers to problem B2 when red is the target colour indicate that both model 2(a) and model 2(b) represent ways in which children handle this problem, 2(a) being rather more common than 2(b). Two subjects in the transition period between stage 1 and stage 2 gave type 2(a) answers, sixteen subjects in stage 2 gave type 2(a) answers, five subjects in stage 2 gave type 2(b) answers and three subjects in the transition period between stage 2 and stage 3 gave type 2(b) answers. This indicates that the developmental sequence may well be from type 1 to type 2(a) to type 2(b) to type 3, but the evidence is scanty and such a conclusion cannot be pressed. Both improvements from type 2 to type 3 and regressions from type 2 to type 1 on problem B2 can be seen in the two transition periods. The same trend can be seen in the secondary school group results in the transition from stage 2 to stage 3 (the secondary school results begin at stage 2, hence they contain no evidence concerning the stage 1 to stage 2 transition).

Cluster analysis of results of the first beads experiment.

The analysis of results as carried out so far is dependent upon the initial interpretations of the children's responses made by the experimenter, albeit in accordance with an explicit classificatory scheme. It has been argued that some form of independent confirmation of the validity of this scheme is necessary, and that high correlation with the interpretations of another scorer is not sufficient for this purpose. TheBrimer cluster analysis, however, does appear to meet this need.

The aim of the Brimer technique is to strip the analysis of as much subjective baggage as possible by making the minimum possible interpretation of the data, which is a decision whether any two responses are the same (identical) or different (in any way). Of course such a decision will always be subjective to a certain extent, for example, most people would not consider the answers 'there are four' and 'there's four there' as being different. Thus complete objectivity is recognised as unattainable from the outset, but its pursuit is not abandoned. The power of the technique depends on the richness and limitations of the available data, rather than the richness or limitations of the experimenter's interpretations. This has certain disadvantages as well as advantages. As Pylyshyn (1972) points out, such methods do not allow the derivation of process-type constructs, since this invariably requires bringing external considerations to bear on the data. 'The end-product of this type of analysis is always some economy-driven description of the variance inherent in a particular sample of observations. The difficulty here is that, as anyone who has ever spent time in an undergraduate laboratory knows, most observational variance is irrelevant to theories or to the understanding of a phenomenon.' (Op.cit., p.191). Fortunately, such criticisms have been circumvented here by the fact that the experimental problems are designed to distinguish various theoretical models, and the fact that the cluster analysis is only being used to add strength to other analyses, not as the sole method of data interpretation.

Although involving minimal interpretation of the data collected, the Brimer analysis cannot be carried out without some theoretical preconceptions, since these enter at the level where decisions concerning the data to be recorded are taken. Most people would not record the colour of the child's sweater, and many more relevant features can be overlooked carelessly or deliberately. (Harris, 1964, argues convincingly that it is impossible to eliminate all preconceptions). The results of the first beads experiment comprise only choice data and verbalisations, other possible data such as latencies or outcomes of choices have been ignored for the purposes of analysis. Any response is ultimately a unique response to a unique situation, and as soon as it is considered in any other way an implicit or explicit theory is involved. In the present experiment the selection of verbal and choice data was dictated by the need for data which would be rich enough to allow the Brimer technique to be profitably used, without being so individual as to lead to a set of almost unique categories.

The starting point of the Brimer analysis is to define an initial set of response categories consisting of the responses given by the first subject to each of the experimental problems. This set of categories is then added to every time a new response to one of the problems is detected until the last response of the last subject has been categorised. This technique was applied to the data generated in the experimental problems B1 - B4, and was carried out with two groups of subjects:

Group A: The primary school children only, i.e. the 72 children aged 5 to 10 years.

Group B: The primary and secondary school children, i.e. the whole sample of 120 children aged 5 to 14 years.

The analysis was performed twice in this way because of the reservations which have been expressed concerning the secondary school children's performance. There is also the additional advantage that the clusters obtained in each analysis can be compared, thus eliminating some of the redundancy inherent in the technique (see Satterly and Brimer, 1971).

The categorisation scheme evolved for the Brimer analysis must now be described. This will be done by considering responses to problem B1 as a typical case and outlining all the categories necessary for their analysis, any extra categories necessary for analysis of responses to the other problems can then be discussed separately.

Problem B1 involves four green and two red beads in one box, and two green and four red beads in the other box. Considering the case where the target colour is green, the following categories of response are distinguishable:

('Pass' choices are defined as choice of 4G2R,'fail' choices are choice of 2G4R, when green is the target colour).

FU: A 'fail' choice, without any reason given, or an unqualified assertion of the rightness of the choice, or repetition of the experimenter's instruction.

e.g. SB51: Chooses 2G4R. 'Don't know why'.
SB3 : Chooses 2G4R. 'To get a green one'.

PU: A 'pass' choice, with a reason like the reasons specified for FU.

e.g. SB4: Chooses 4G2R. 'That's my best one'.

FSR: A 'fail' choice with a spurious reason, e.g. the colour of the box.

e.g. SB5: Chooses 2G4R. 'Got brown on it'. (The box containing the beads has brown on the side).

PSR: A 'pass' choice with a reason like FSR.

e.g. SB8: Chooses 4G2R: 'I like brown'.

FQS: A'fail' choice, accompanied by implicit or explicit attempts to quantify spurious factors.

e.g. SB10: Chooses 2G4R. 'Got more green round the corners of the box.'

PQS: A 'pass' choice with a reason like FQS.

e.g. SB14: Chooses 4G2R. 'Same again. It's always best, always has more patches on it'.

PPR: A 'pass' choice with a reason based on previous outcomes or previous choices, e.g. left-right alternation or win-stay lose-shift strategy.

e.g. SB37: Chooses 4G2R. 'I got green last time'.

PRF: A 'pass' choice accompanied by reference to some relevant factor not included in the other categories.

e.g. SB9: Chooses 4G2R. 'It's got a green one in it'.

FQA: A 'fail' choice accompanied by reference to the number of beads of the target colour in the box chosen.

e.g. SB13: Chooses 2G4R. 'Two greens in here'.

PQA: A 'pass' choice accompanied by a reason like FQA.

e.g. SB17: chooses 4G2R. 'It's got four greens'.

PQTM: A 'pass' choice accompanied by a claim that there are 'more' beads of the target colour in the box chosen than in the other box, without specification of numbers.

e.g. SB30: Chooses 4G2R 'More greens than in that one'.
SB39: Chooses 4G2R 'The most greens'.

PQTA: A 'pass' choice accompanied by a reason referring to the number of beads of the target colour in each box, and choice of the box with more'.

e.g. SB29: Chooses 4G2R. 'Got four in; other has two'.

PQTR: A 'pass' choice accompanied by specification of the ratio of the number of target beads in one box to the number of target beads in the other box.

e.g. SB63: Chooses 4G2R. 'Twice as many greens as the other one'.

PQNTL: A 'pass' choice accompanied by reference to the non-target beads in each box, and the fact that the box chosen has 'less'.

e.g. SB54: Chooses 4G2R. 'Less red than the other box'.

PQRA: A'pass' choice based on comparison of the numbers of target and non-target beads in each box.

e.g. SB48: Chooses 4G2R. 'Four red and only two green in the other box. Also only two red in this box and four green'.

PQROA: A 'pass' choice accompanied by specification of the number of target and non-target beads in the box not chosen.

e.g. SB46: Chooses 4G2R. 'Only two greens in the other and four reds'.

PQRM: A 'pass' choice accompanied by a claim that there are 'more' beads of the target colour than beads of the non-target colour in the box chosen, and that the situation in this respect is more favourable in the box chosen than the box not chosen.

e.g. SB47: Chooses 4G2R. 'More greens than reds. Other has more reds than greens'.

PQRS: A 'pass' choice accompanied by a statement of how many more target than non-target beads are in the boxes.

e.g. SB67: Chooses 4G2R. 'More chance again. Two more greens than reds. Other is two less'.

The following categories were required,in addition to those listed previously, in order to analyse the secondary school results together with the primary school results.

PQRR: A 'pass' choice accompanied by specification of the ratio of target to non-target beads in the box chosen.

e.g. SB108: Chooses 4G2R. 'There's four greens and two reds. A two-to-one chance.'

U: Refusal to choose, together with 'philosophical' justification.

e.g. SB97: Refuses to choose. 'Either one. It doesn't matter, you can get red or green every time'.

This completes the list of categories necessary to classify the responses obtained to the section of problem B1 involving green as the target colour. For the section of problem B1 involving red as the target colour a new set of categories was drawn up. Many of these were identical to those already specified (except that they are new categories because this is a new problem) and will not be mentioned. The following additional categories were necessary.

PQANT: A 'pass' choice (which now becomes choice of 2G4R) accompanied by a reference to the number of non-target beads in the box chosen.

e.g. SB69: Chooses 2G4R. 'Only two greens in there'.

PQTS: A 'pass' choice accompanied by specification of how many more target beads are in the box chosen than in the box not chosen.

e.g. SB87; Chooses 2G4R. 'There's two more reds than there'.

PQRL: A 'pass' choice accompanied by a claim that there are 'less' beads of the non-target colour than beads of the target colour in the box chosen.

e.g. SB54: Chooses 2G4R. 'Less green than red ones'.

Problem B2 has four green beads and three red beads in one box, and two green beads and three red beads in the other box. When the target colour is green, only two new categories are required in addition to others similar to those already described.

FRF: A'fail' choice (choice of 2G3R) accompanied by reference to some relevant factor not included in the other categories.

e.g. SB9: Chooses 2G3R. 'That one's got a green one in'.

FQTL: A 'fail' choice accompanied by the reason that there are 'less' of the target beads in the box chosen.

e.g. SB55: Chooses 2G3R. 'Only two greens. Could get green easily. Less green than in that box.'

When red is the target colour in problem B2, there are three reds in each box. A pass choice is defined as choice of 2G3R, and choice of 4G3R or assertions that either box may be chosen are considered fail choices. The categorisation scheme must be expanded to accomodate this possibility, and in addition to the categories already described, the following must be introduced:

FPR: A 'fail' choice with a reason referring to previous choices or outcomes, e.g. left-right alternations, or win-stay lose-shift strategies.

e.g. SB37: Chooses 4G3R. 'Same again. Was red before, could be again'.

FQTA: A 'fail' choice or refusal to choose following an assertion that both boxes have the same chance of success because there are three reds in each. (This is slightly different to the way FQTA is used as a category in later problems).

e.g. SB28: Chooses 4G3R. 'Don't know why. Both got three in. Doesn't matter which one. This has mostest greens'.

FQTM: A 'fail' choice or refusal to choose following an assertion that both boxes have the same chance of success, without any further reason. (This is slightly different to the way FQTM is used as a category in later problems).

e.g. SB41: Chooses 4G3R. 'Just a guess. Same really. I think'.

SB62: 'They're both the same. I'll pick that one.' Chooses 4G3R.

PQTA: A 'pass' choice following an assertion that both boxes have the same chance of success because there are three reds in each. (This is slightly different to the way PQTA is used as a category in the other problems).

e.g. SB31: Chooses 2G3R. 'Can take any one. Both got three in. Guess'.

PQTM: A 'pass' choice following an assertion that both boxes have the same chance of success, without any further reason. (This is slightly different to the way PQTM is used as a category in the other problems).

e.g. SB51: Chooses 2G3R. 'Don't know why. Just guessed. Any will do'.

PQNTA: A 'pass' choice with an assertion that both boxes have the same number of target beads. The reason for the choice refers to the number of non-target beads in the box chosen.

e.g. SB63: Chooses 2G3R. 'Only two greens. Same reds in each one'.

FQRM: A 'fail' choice accompanied by the reason that there are 'less' target beads than non-target beads in the box chosen.

e.g. SB55: Chooses 4G3R. 'Not as many reds as greens, easily get red'.

A further category is necessary for the secondary school results:

FQNTM: A 'fail' choice following an assertion that both boxes have an equal chance of success. The reason given for the choice is that there are 'more' non-target beads in the box chosen than in the other box.

e.g. SB85: Chooses 4G3R. 'Both got three. I'll have that one because there's more greens'.

Problem B3 has four green beads and four red beads in one box, and two green beads and two red beads in the other box. This means that when the target colour is green or red, a pass response will involve the assertion that both boxes have the same chance of success, and failure to make this assertion will be considered a fail. In order to modify the categories generated so far to make them useful in this situation, the suffixes L and M can be added to the 'fail' categories. These suffixes indicate choice of the box with less or with more beads. Thus category FU can become FUL or FUM in this situation, the reason in each case being the same, but the choice different. If the categorisation scheme is modified in this way many of the responses to problem B3 can be handled satisfactorily, but some further categories are still necessary. When green is the target colour, the following additional categories must be introduced:

FMM: A 'fail' choice of the box with more beads of both colours, accompanied by an assertion to the effect that this is the reason for the choice.

e.g. SB34: Chooses 4G4R. 'More greens and more reds'.

FQTA: FQTM: FQTS: FQTR: These categories are the same as the categories PQTA: PQTM: PQTS: PQTR: which were outlined for problem 1, except that in this problem the reason involved lead to'fail'choices rather than'pass'choices.

Only one further category is now needed for the results of problem B3 when red is the target colour:

PU: A 'pass' assertion with a 'don't know why' reason.

e.g. SB66: Chooses 4G4R. 'They're both equal. Don't know really'.

Problem B4 has six green beads and four red beads in one box, and two green beads and two red beads in the other box. For the part involving green as the target colour no additional categories were required for the primary school results, and only one for the secondary school results:

FQNTL: A 'fail' choice with the reason that the box chosen has not so many non-target beads as the other box.

e.g. SB115: Chooses 2G2R. 'You can easily pick a red in the other one'.

For the part of problem B4 involving red as the target colour, the following additional categories are necessary: (Pass choices are choice of 2G2R, fail is choice of 6G4R).

PQRF: A 'pass' choice accompanied by quantification of some relevant factor not included in the other categories.

e.g. SB16: Chooses 2G2R. 'Got four in'.

PLL: A 'pass' choice accompanied by the reason that there are less target and non-target beads in the box chosen.

e.g. SB35: Chooses 2G2R. 'There's less greens and less reds'.

FQANT: A 'fail' choice accompanied by specification of the number of non-target beads in the box chosen.

e.g. SB15: Chooses 6G4R 'Got six green ones.'

PQTL: A 'pass' choice with the reason that there are less beads of the target colour in the box chosen than in the other box.

e.g. SB20: Chooses 2G2R. 'Got lessest red'.

FQRA: A 'fail' choice accompanied by specification of the number of target and non-target beads in the box chosen.

e.g. SB34: Chooses 6G4R. 'There's four reds and six greens'.

This completes the description of the categorisation scheme evolved for the cluster analysis. The result of applying this scheme to the experimental data can be found in Appendix C. When the primary school results only are analysed there are 72 subjects, each giving 8 responses, and 141 categories are required. For the combined primary and secondary school results there are 120 subjects, each giving 8 responses, and 164 categories are required. Each category was then given a separate number and the data, in the form of a list of eight numbers for each subject (the categories corresponding to his responses), was analysed by computer.

Unfortunately the computer's maximum capacity was 138 categories when there were 72 subjects, and 130 categores when there were 120 subjects. This means that the number of categories must be reduced by 3 for the primary school results, and by 34 for the combined primary and secondary school results. The simplification of the primary school categories was effected by including PQRL in PQRM in the second part of problem 1, and by including PQROA in PQRA in both parts of problem 1. As there was only one example of each of these categories, this cannot be expected to affect the result of the cluster analysis significantly.

The simplification of the combined primary and secondary school results is more difficult. This analysis is intended to provide a more general overview of the whole result of the experiment than the primary school analysis, so the decision was taken to 'tidy up' the categories which one might describe as poor performance categories, i.e. the large number of small categories produced by the five and six year olds. This is more desirable than impoverishing data from the older children, and the categories produced by the younger children have already been analysed in the primary school only analysis.

The simplification which was arrived at, involved treating the following groups of categories as single categories:

Problem B1, green target:	(PSR,PQS,PPR), (FSR,FQS), (PU,PRF),(PQRA,PQROA).
Problem B1, red target:	(PSR,PQS,PPR),(FSR,FQS),(PQRA,PQROA),(PQRL,PQRM).
Problem B2, green target:	(PSR,PQS),(FU,FRF),(PU,PRF).
Problem B2, red target:	(PSR,PQS,PPR),(FSR,FQS,FPR),(PU,PRF).
Problem B3, green target:	(FSRL,FQSL,FPRL),(FSRM,FPRM).
Problem B3, red target:	(FSRL,FQSL),(FSRM,FQSM,FPRM),(FUL,FRFL).

Problem B4, green target: (PSR,PQS,),(FSR,FQS,FPR),(FU,FRF).
Problem B4, red target: (PSR,PQS),(FSR,FQS,FPR),(FU,FRF,FQANT),(PU,PQRF).

This removes 35 categories from the analysis, leaving 129 categories. The list of category numbers for each subject in the two analyses can be found in Appendix C.

The cluster analysis program generates groups of subjects who exhibit covarying categories. The groups are generated according to the criterion of a chi-square value significant at the 0.05 level. Second-order groups (clusters) of the original groups are then generated, again using 0.05 as the level of significance. The purpose of this second application of the technique is to reduce the high redundancy which usually occurs in the first-order groupings. A more complete and formal description of the method can be found in Appendix I.

Results of the cluster analysis of data from the primary school children.

The results of the two cluster analyses will be discussed in turn, starting with the analysis of the primary school results only. This analysis generated 16 first-order groups, and 3 second-order groups, or clusters. The detailed composition of these groups and clusters can be found in Appendix C. The groupings are arrived at on a purely statistical basis, so an attempt will be made to expose the factors underlying them. Enough information is given in Appendix C for the interested reader to compare his interpretations with the writer's.

Group 1: A group of subjects who seem to solve the problems by seeing which collection of beads contains more of the target-colour beads, but who do not use the word 'more' in their answers, and refer only to the number of target-colour beads in each collection. The group has 6 members.

Group 2: A group of subjects who seem to solve the problems by seeing which collection of beads contains more of the target-colour beads, and who use the word 'more' in their answers, usually without mentioning any numbers. The group has 28 members.

Group 3: A group of subjects who seem to solve the problems by comparing the relative number of target-colour beads and non-target colour beads in each collection, and use the word 'more' in their answers. The group has 17 members.

Group 4: A group of subjects who make choices but don't give reasons, or whose reasons are repetitions of the experimental instructions or unqualified assertions of the correctness of the choices. The choices made by these children look erratic. The group has five members.

Group 6: A similar group to group 3.

Group 3 (17 members) and group 6 (14 members) have 12 members in common.

Group 7: Similar to group 5.

Group 5 (5 members) and group 7 (5 members) have 3 members in common.

Group 8: A group of subjects whose answers indicate quantifications of numbers of target beads, but do not always refer to both collections. The choices made by subjects in this group show an overall consistency with choices made by subjects in group 1 or group 2, but some individual items deviate from this pattern. The group has 4 members.

Group 9: Similar to group 2.

Group 2 (28 members) and group 9 (29 members) have 27 members in common.

Group 10: Similar to group 3 and group 6.

Group 3 (17 members) and group 10 (13 members) have 12 members in common.

Group 6 (14 members) and group 10 have 6 members in common.

Group 11: A group of subjects who seem to solve the problems by comparing the relative number of target-colour beads and non-target colour beads in each collection, but who do not use the word 'more' in their answers, and refer only to the number of target-colour beads and non-target-colour beads in each collection. The group has 5 members.

Group 12: Similar to group 4.

Group 4 (6 members) and group 12 (5 members) have 5 members in common.

Group 13: Similar to group 4 and group 12.

Group 4 (6 members) and group 13 (5 members) have 4 members in common.

Group 12 (6 members) and group 13 have 3 members in common.

Group 14: A group of subjects whose answers involve attempts to quantify factors which seem irrelevant to adults. Choices made by these subjects look erratic. The group has 5 members.

Group 15: Similar to group 14.

Group 14 (5 members) and group 15 (4 members) have 3 members in common.

Group 16: Similar to group 4, group 13, and group 12.

Group 4 (6 members) and group 16 (3 members) have 2 members in common.

Group 13 (5 members)and group 16 have 2 members in common.
Group 12 (6 members) and group 16 have 1 member in common.

If the preceding interpretation of the groups generated is valid, four groups of groups, or clusters, might be expected:

Cluster A: A cluster of groups of subjects who make erratic choices and don't really justify their choices. This cluster will consist of group 4, group 12, group 13, and group 16.

Cluster B: A cluster of groups of subjects who choose erratically and give irrelevant reasons (group 5 and group 7) or give reasons involving quantification of irrelevant factors (group 14 and group 15). This cluster will consist of group 5, group 7, group 14, and group 15.

Cluster C: A cluster of groups of subjects who seem to consistently choose the box with the larger number of beads of the target-colour, and whose answers involve reference to the number of target-colour beads in the box chosen (group 8), both boxes (group 1), or both boxes with the term 'more' substituted for exact numbers (group 2 and group 9). This cluster will consist of group 1, group 2, group 8, and group 9.

Cluster D: A cluster of groups of subjects who seem to consistently choose the box with the most favourable relative number of target-colour beads and non-target beads. Answers given by these subjects involve reference to numbers of target and non-target beads (group 11), or to numbers implied by use of the term 'more' (group 3, group 6, group 10). This cluster will consist of group 3, group 6, group 10, and group 11.

These predicted clusters are similar in outline to the types of response to the experiment predicted from theoretical considerations. Clusters A and B correspond to model 1 (Piaget's first developmental stage), cluster C corresponds to model 2 (Piaget's second developmental stage) and cluster D corresponds to model 3 (not reported by Piaget). The actual clusters generated by the cluster analysis are as follows:

First cluster:	Group 7
	Group 14
	Group 15
	Group 5
Second cluster:	Group 13
	Group 16
	Group 4
	Group 12

Table 3: Comparison of stage groupings of first board experiment subjects from interpretative and cluster analyses.

Cluster analysis groupings

Subjects in stage 1 groups only:

SD1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 11, 12, 14,
25, 26,
37,
49.

Subjects in stage 1 and stage 2 groups, i.e. first transition:

SD16,
27,
50.

Subjects in stage 2 groups only:

SD13, 15, 17, 18, 19,
20, 21, [illegible], [illegible], 24, [illegible], [illegible],
30, 31, [illegible], 33, [illegible], 39,
40, 41, 42, 43, 44,
51, 52, 53,
61, 62, 64.

Subjects in stage 2 and stage 3 groups, i.e. second transition:

SD 36,
46,
56, 58,
65,
71.

Subjects in stage 3 groups only:

SD [illegible], 35,
45, 47, 48,
54, 55, 57, 59,
[illegible], [illegible], 67, [illegible], 69,
7[illegible], 70

Subjects in stage 1 and stage 3 groups:

None.

Interpretative analysis groupings.

Stage 1:

SD1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 12,
25, 26,
49.

Transition 1 - 2:

SD11, 14, 15, 16, 17, 18, 19,
20, 27,
37,
50, 51.

Stage 2:

SD13,
21, [illegible], [illegible], 28, 29,
30, 31, 32, 33, 38, 39,
[illegible], 41, 42, 43,
52, 53,
61, 62, 63.

Transition 2 - 3:

SD[illegible],
[illegible], [illegible], 3[illegible],
[illegible], [illegible], [illegible], 4[illegible],
54, 55, 56, 57, 58, 59,
60, 64, 65, 66,
71.

Stage 3:

SD [illegible],
[illegible], [illegible], [illegible],
7[illegible], 72.

Third cluster: Group 3
Group 6
Group 10
Group 2
Group 9
Group 1
Group 11
Group 8.

(The groups in each cluster are listed in the order printed out by the program, corresponding to decreasing order of weighting to the cluster). In other words the first cluster is predicted cluster B, the second cluster is predicted cluster A, and the third cluster consists of predicted cluster C combined with predicted cluster D.

Discussion: This result provides broad support for the proposed stage description of the results. The three developmental stages, corresponding to models 1, 2, and 3 show up clearly in the initial groups, and in the clusters models 2 and 3, which correspond to systematic ways of dealing with the experimental problems, are separated from the more erratic strategies. The fact that two versions of model 1 are distinguished, both in the groups and the clusters, suggests that something important may have been overlooked.

The failure of models 2 and 3 to appear separately in the clusters is initially disappointing, but might have been expected because of the large overlap, in the form of a transition, between the developmental stages these models represent (see table 2). This can be seen clearly if the first-order groups corresponding to the proposed developmental stages are examined. As already indicated: stage/model 1 seems to be represented by groups 4, 12, 13, and 16, together with groups 5, 7, 14, and 15.

Stage/model 2 corresponds to groups 1, 2, 8, and 9.
Stage/model 3 corresponds to groups 3, 6, 10, and 11.

If there is a developmental sequence between these stages of the kind indicated in table 2, we would expect to find an overlap between some subjects in the groups corresponding to stage 1 and stage 2, and some overlap between subjects in the groups corresponding to stage 2 and stage 3, but no overlap between subjects in the groups corresponding to stage 1 and stage 3. It would also be expected that the stage groupings of subjects will be similar to those in table 2.

In order to test this possibility, table 3 was compiled. This table shows the stage groupings and transitions derived from the cluster analysis (by the method outlined) alongside the stages and and transitions outlined in table 2. The most noticeable difference is that in the cluster analysis the stages are larger and the transitions smaller than in the 'interpretative' analysis. The reason for this is undoubtedly that the criterion of statistical association used in the cluster analysis is less severe (although more objective) than the criteria drawn up for interpretation by the experimenter (which insist that unless all responses correspond to the same stage the subject will be considered as transitional). This causes the difference between the cluster analysis groupings and the interpretative groupings that many subjects, placed in transition by the interpretative procedure, are shifted into adjoining stages by the cluster analysis.

There is one other difference between the cluster analysis stage groupings and the interpretative groupings, namely that SB63 is omitted from the cluster analysis groups entirely. The reason for this is the unique responses given by SB63, many of which were of the form 'there's twice as many reds/greens in this box as the other box'. Such use of ratios with a model 2 kind of overall strategy is unique in the primary school analysis, so that SB63, who was thought by the experimenter to belong in stage 2, cannot be grouped statistically with any other subjects, since he has no responses in common with any other subject.

The cluster analysis, then, can be considered as supporting the stage description proposed earlier, but it also goes further because it distinguishes two versions of stage 1. These two versions of stage 1 are represented by the first cluster and the second cluster. One of these clusters seems to consist of subjects who don't justify their choices, and the other consists of subjects who give irrelevant (by adult standards) justifications for their choices. It is possible that these two versions of model 1 may also fit into the developmental sequence, and for this reason the overlap between the subjects in the two clusters and subjects in other clusters must be investigated.

The subjects in the groups of the first cluster (erratic choices and irrelevant reasons) are: SB 2,5,8,10,13,14,25,26,27,37,49.

The subjects in the groups of the second cluster (erratic choices without justification) are: SB 1,3,4,6,7,9,11,16,50.

These clusters of subjects do not overlap with each other at all, but both overlap withsubjects in the first transition as defined by both the cluster analysis and the experimenter's interpretative analysis (see table 3). This suggests that they represent parallel but separate developments.

Hence, instead of the developmental model,

Stage 1 $\rightarrow$ Stage 2 $\rightarrow$ Stage 3,

we have the model,

Stage 1A $\searrow$
Stage 1B $\rightarrow$ Stage 2 $\rightarrow$ Stage 3.

At present, however, this remains little more than a possibility, because the impoverished nature of the verbal responses of children at the stage 1 level does not permit satisfying inferences from verbal data, It is quite possible that stage 1A and stage 1B are merely artifacts of this impoverishment, and a different method of investigation would have to be devised to follow this up.

The overall conclusion to be derived from the cluster analysis of the primary school results is that the general developmental stage analysis, stage/model 1 to stage/model 2 to stage/model 3 is confirmed. Model 4 does not show up in the results, but then only children up to age 10 are being considered. The model 2(a) or model 2(b) distinction is also not apparent in the cluster analysis, but this might be expected because only one part of one experimental problem is sensitive to this distinction. The possibility that there are two main types of strategy rather than a unitary stage 1 strategy is opened up. This does not contradict anything said previously as no model for stage 1 has been proposed, but the evidence for the distinction is equivocal, and may be an artifact of the impoverishment of stage 1 verbal responses.

<u>Results of the cluster analysis of data from the combined primary and secondary school groups</u>.

The analysis of the combined primary and secondary school results involves 120 subjects, whose responses are grouped into 129 categories. Initially, there were 164 categories, but 35 were removed in the manner already described, in order to assimilate the data to the handling capacity of the computer. This simplification was effected with the data from the stage 1 subjects, so that it will not affect the overall analysis, as the stage 1 data have already been analysed by the primary school only cluster analysis, and the response categories produced by most of the primary school children and all the secondary school children are not altered in any way. It does mean, however, that the distinction between the two types of stage 1 response noted earlier should disappear.

The cluster analysis of the combined primary and secondary school results generated 18 first order groups and 3 clusters. The composition of the groups can be found in Appendix C and an interpretation of the factors underlying each group will be offered here.

Group 1: A group of subjects who seem to solve the problems by seeing which collection of beads contains more beads of the target colour and who use the word 'more' in their answers. The group has 42 members.

Group 2: A group of subjects who seem to solve the problems by comparing the relative number of target-colour beads and non-target-colour beads in each collection, and use the word 'more' in their answers. The group has 41 members.

Group 3: A group of subjects who choose inconsistently and offer reasons for their choices which are irrelevant by adult standards. (The variety of response categories in this group has been reduced in the preliminary stages of the analysis). The group has 9 members.

Group 4: A group of subjects who seem to solve the problems by seeing which collection of beads contains more of the target-colour beads, but who do not use the word 'more' in their answers, and refer only to the number of target -colour beads in each collection. The group has 9 members.

Group 5: Similar to group 2.

Group 2 (41 members) and group 5 (14 members) have 13 members in common. The difference between group 5 and group 2 seems to be that members of group 5 manage to apply the characteristic 'group strategy' to problem B3, whereas members of group 2 often don't manage this.

Group 6: Similar to group 2 and group 5.

Group 2 (41 members) and group 6 (27 members) have 21 members in common.

Group 5 (14 members) and group have 13 members in common.

Group 7: A group of subjects who choose inconsistently and offer no reasons for their choices, or mention relevant factors in a haphazard way. (The variety of response categories in this group has been reduced in the preliminary stages of the analysis). The group has 7 members.

Group 8: Similar to group 3.

Group 3 (9 members) and group 8 (10 members) have 7 members in common.

Group 9: Similar to group 2, group5, and group 6.

Group 2 (41 members) and group 9 (40 members) have 38 members in common.

Group 5 (14 members) and group 9 have 14 members in common, i.e. group 5 is a more tightly defined subgroup of group 9.

Group 6 (27 members) and group 9 have 21 members in common.

Group 10: Similar to group 7.

Group 7 (7 members) and group 10 (6 members) have 5 members in common.

Group 11: Similar to group 1.

Group 1 (42 members) and group 11 (39 members) have 37 members in common.

Group 12: A group of subjects who consistently seem to solve the problems by seeing which collection of beads contains more beads of the target colour, but who don't use the word 'more' in their answers, and often only refer to the number of target-colour beads in one of the two boxes. The group has 6 members.

Group 13: Similar to group 1 and group 11.

Group 1 (42 members) and group 13 (42 members) have 41 members in common.

Group 11 (39 members) and group 13 have 36 members in common.

Group 14: Similar to group 2, group 5, group 6, and group 9.

Group 2 (41 members) and group 14 (38 members) have 34 members in common.

Group 5 (14 members) and group 14 have 14 members in common, i.e. group 5 is a more tightly defined subgroup of group 14.

Group 6 (27 members) and group 14 have 18 members in common.

Group 9 (40 members) and group 14 have 36 members in common.

Group 15: A group of subjects who seem to solve the problems by comparing the relative number of target-colour beads and non-target-colour beads in each collection, but who don't always use the word 'more' in their answers, often referring only to numbers of target-colour beads and non-target colour beads in each collection. The group has 8 members.

Group 16: Similar to group 2, group5, group6, group 9, and group 14.

Group 2 (41 members) and group 16 (26 members) have 20 members in common.

Group 5 (14 members) and group 16 have 11 members in common.

Group 6 (27 members) and group 16 have 18 members in common.

Group 9 (40 members) and group 16 have 19 members in common.

Group 14 (38 members) and group 16 have 15 members in common.

Group 17: Similar to group 7 and group 10.

Group 7 (7 members) and group 17 (7 members) have 6 members in common.

Group 10 (6 members) and group 17 have 4 members in common.

Group 18: A curious group of two subjects who don't seem to 'belong' together, except that both subjects would be classed as Stage 2 in the 'interpretative' analysis.

These initial groupings are similar to those identified in the cluster analysis carried out on the primary school results only. This confirms the view that the data obtained from the results of the experiment with the secondary school children fall into a similar pattern to the primary school results. The combination of the primary and secondary school results, with the consequent increased frequency of the more 'advanced' categories, leads to a more detailed analysis of relations between the more 'advanced' categories whose result can be seen in the large number of overlapping groups. The fact that so much overlap occurs between certain sets of groups, is, of course, evidence for common factors underlying these groups. There is still no sign of a group of the categories corresponding to model 4 (the 'true proportion' model, Piaget's third developmental stage) although certain individual categories do appear to be something of the kind.

What clusters might be expected on this basis? Probably something similar to the kind of clusters which were anticipated in the first cluster analysis:

Cluster A: A cluster of groups of subjects who choose erratically and either don't really justify their choices (group 7, group 10, group 17) or give justifications which are irrelevant by adult standards (group 3, group 8). This cluster will consist of group 3, group 7, group 8, group 10 and group 17.

Cluster B: A cluster of groups of subjects who seem to consistently choose the box with the larger number of beads of the target-colour, and whose answers involve reference to the number of target-colour beads in the box chosen (group 12), both boxes (group 4), or both boxes with the term 'more' substituted for exact numbers (group 1, group 11, group 13). Also the 'odd' group 18. This cluster will consist of groups 1, group 4, group 11, group 12, group 13, and group 18.

Cluster C: A cluster of groups of subjects who seem to consistently choose the box with the most favourable relative number of target-colour beads and non-target-colour beads. Answers given by these subjects involve reference to numbers of target and non-target beads (group 15), or to numbers implied by use of the term 'more' (group 2, group 5, group 6, group 9, group 14, group 16). This cluster will consist of group 2, group 5, group 6, group 9, group 14, group 15 and group 16.

These predicted clusters correspond to developmental stages 1, 2, and 3 respectively. The split between the two varieties of stage 1 response evident in the primary school cluster analysis is not anticipated here because of the artificial way in which the homogeneity of the stage 1 responses was increased by the elimination of certain categories.

The actual clusters produced by the cluster analysis are as follows:

First cluster	Group 3
	Group 10
	Group 8
	Group 7
	Group 17
Second cluster	Group 11
	Group 18
	Group 12
	Group 1
	Group 13
	Group 17
	Group 7
	Group 8
Third cluster	Group 5
	Group 9
	Group 14
	Group 6
	Group 2
	Group 16
	Group 15
	Group 4
	Group 1
	Group 13.

(The groups in each cluster are listed in the order generated by the program).

This result is best described by considering in turn both the groups in each cluster, and the groups not in each cluster.

The first cluster consists of the predicted cluster A, in other words all the groups which look like the results of stage 1 strategies. All stage 2 groups and stage 3 groups are excluded from the first cluster.

The second cluster consists of all except one of the stage 2 groups, that is all except one group from the predicted cluster B, plus three stage 1 groups, which are also in the first cluster. The groups excluded from the second cluster are all the stage 3 groups, two of the stage 1 groups, and one of the stage 2 groups. Inspection of the weightings of all the individual groups to this cluster reveals however, that the exclusion of the stage 2 group (group 4) is because its weighting is positive but fails to meet the criterion of statistical significance. What has happened is that group 4 shows a much stronger weighting to the second cluster than the other excluded groups, but this weighting is not quite strong enough to be significant.

The third cluster consists of all the stage 3 groups and three of the stage two groups (including group 4). The groups excluded from this cluster are all the stage 1 groups and three of the stage 2 groups. Again, inspection of the weightings shows that the three excluded stage 2 groups are weighted positively towards this cluster, but their weightings are not significant. The weightings of the excluded stage 1 groups are, by contrast, all negative or almost zero.

Discussion: Once again the clustering section of the analysis has separated the erratic strategy represented by stage 1 from the strategies of stage 2 and stage 3. In addition the developmental sequence from stage 1 to stage 2 to stage 3 is supported because the clusters are made up of groups corresponding to stage 1, stage 1 and stage 2, stage 2 and stage 3, respectively. In other words, the clusters separate stage 1 and stage 3, but mix stage 1 with stage 2, and stage 2 with stage 3. This is what would be expected from a three-stage developmental sequence with two transitions, because the statistical association criterion used in the analysis is not as strict as the criteria used in the initial interpretative analysis. Groups corresponding to stage 1, stage 2 and stage 3 show up clearly in the initial groupings, so that independent confirmation of the original interpretation of the experiment is provided.

Summary of result of the first beads experiment.

This concludes the description of the Brimer cluster analysis and brings the description of the first beads experiment to a point where it can be summed up by restating the aims initially put forward for the experiment, together with the conclusions that can be derived from the results of the experiment.

The first aim of the experiment was:

A. To find out whether the models put forward provide satisfactory models of children's behaviour when faced with this type of problem.

The conclusions regarding this aim are dependant on the way the aim is initially interpreted. It is clear from the analysis of the results of the experiment that the proposed models are satisfactory descriptions of the overall strategies employed by children in the solution of problems of this type. It is equally clear that they do not account for much of the variance in the children's behaviour, as is shown by the large number of first order groups corresponding to each of the three main strategies in the two cluster analyses. There are many possible reasons for this. The models may need refinements of detail in order to account for the varieties of individual usage, or some other factor may be involved. It is possible, for example, that the models outlined describe accurately the strategies used by individual children to handle the information contained in these kinds of problems, but that an additional set of models is required to illustrate the way children initially organise, or 'code' this information. This possibility will be pursued in some detail later. Reservations must also be expressed about the stage 1 strategy for which no actual model has been proposed, and which may well be a ragbag of different strategies. This possibility is supported by the result of the first cluster analysis, which distinguished two varieties of stage 1, but this may have been an artifact of the impoverishment evident in the stage 1 verbal responses.

The second aim of the experiment was:

B. To find out whether model 2(a) or, model 2(b) provides the more accurate description of Piaget's second developmental stage, or whether both apply.

The experimental results leave little doubt that model 2, or Piaget's second developmental stage, is a valid description of a strategy employed by many children to solve this kind of problem. The results of the 'interpretative' analysis indicate that both versions of model 2, i.e. 2(a) and 2(b), are used by different children in the 'tricky' case when there are equal numbers of target-colour beads in the two collections of beads. The cluster analysis, however, does not seem to distinguish children using 2(a) and children using 2(b), because only one item out of the eight involved in the experiment (four problems, each in two parts)was sensitive in this respect.

The third aim of the experiment was:

C. To find out whether model 3 is only a theoretical possibility, or if it corresponds to an actual strategy employed by children to solve problems of this type.

The result of the experiment supports the view that, in the present type of situation at least, model 3 does represent a strategy employed by many children. Support is also leant to the contention that model 4 corresponds to a strategy which is not common in the sample investigated, and that previous researchers may have confused the strategies represented by models 3 and 4. This possibility was mentioned as a consideration in the design of the experiment, but its investigation was not one of the explicit aims of the experiment. The results also show little evidence of anything like the combinatorial system of Piaget's formal operational stage, but this cannnt be stated with certainty without a sample of even older subjects. This would be impracticable because of the adverse reactions to the experiment by secondary school children which have been mentioned. The final aim of the experiment was:

D. To find out whether there is a developmental sequence from the first stage described by Piaget to model 2 to model 3 to model 4, or even a sequence from Piaget's first stage to model 2(a) to model 2(b) to model 3 to model 4.

Throughout the analysis a general developmental sequence from Piaget's first stage (stage 1) to a strategy corresponding to model 2 (stage 2) to a strategy corresponding to model 3 (stage 3) is found. No evidence of further development to a stage 4 corresponding to the model 4 strategy was produced, although some subjects did produce solutions of the model 4 type. The possibility of a stage 4 cannot be completely ruled out, as the performance of the secondary school group was poor by comparison with that of the primary school group, and the maximum age of the subjects was 14 years, which is arguably not far enough into Piaget's formal operational period for an adequate test. In addition a study by Fischbein et.al (1970), which seems to satisfactorily take account of the model 3 strategy, does find a 'true proportion' stage, and in particular find that many children aged 9 - 10 and 12 - 13 will give answers based on comparison of proportions after a short instruction. This instruction involved showing how to use a grouping technique to answer questions, so that the procedure adopted is to consider, say, the number of red beads for every green bead in each collection. A similar strategy was found to occur spontaneously in a different situation (making 'walls' of the same length with different Cuisenaire rods) by Lunzer and Pumfrey (1966).

Evidence that model 2(a) and model 2(b) also represent steps in this developmental sequence was produced in the 'interpretative' analysis, but was not confirmed by the cluster analysis. The cluster analysis of the primary school results did, however, reveal the possibility that stage 1 describes two separate strategies used by children, and that these strategies develop independently into stage 2, with no children exhibiting both forms of stage 1. This remains to be confirmed by a more suitable technique for use with five and six year olds.

Some possible criticismsand a control experiment.

Now that the conclusions to be drawn from the first beads experiment have been summarised, there are certain common criticisms of this type of experiment which must be considered. One such criticism is that too much stress is placed on verbal reports by children of what they are doing, and the analysis of these reports, throughout the experiment. This criticism was discussed in some detail during the preliminary stages of the design of the experiment, and it was argued that use of verbal data could be justified if the following conditions were met:

(a) The experimental design is arranged to 'disambiguate' the reasons for their choices given by the subjects as much as possible.

(b) The choices made by the subjects and their justifications of these choices do not conflict with each other.

(c) The analysis of the verbal data involves a minimal amount of interpretation by the experimenter.

The design, results and analysis of the experiment appear to have satisfied these conditions.

It is possible to criticise the lack of full 'randomisation' in the experimental design. One or two cavalier measures were taken in this respect in order to speed up the execution of the experiment and remove artificial holdups of the kind where the experimenter must consult his instructions to find out the exact specification of the next trial. Such measures included shuffling the experimental cards betbetween problems rather than following a latin square ordering, and always using the same set of cards instead of mixing two sets, one set having a red-green arrangement. The consistency of the experimental results and the lack of evidence of colour preferences indicates that the randomising measures which were taken were sufficient, and there is little danger of artifacts arising from order-effects.

Another possible artifact might by caused by the reinforcement of successful choices, either by direct pay-off or experimenter's approval. The direct pay-off type of reinforcement does not seem to affect the choices of any except stage 1 subjects, as the consistency of the stage 2 and stage 3 choices demonstrates, but the possibility of unwitting reinforcement by the experimenter himself of the type of answers he likes can never by ruled out, even though every effort was made to avoid it.

It can be argued that in an experiment of this type, involving a number of similar trials, the subjects' responses may show a spurious consistency, inthe the sence that many subjects will choose a way of getting an answer, and then stick to it. Effects like this cannot be eliminated in any experiment involving more than one trial, although the present experiment minimises them by explicitly trying to dislodge subjects from their preferred strategies with 'tricky' problems. Nevertheless, the results can be interpreted as supporting a weak form of this criticism.

A more sophisticated criticism, which is in some respects similar to the preceding criticism, is that the solution strategies found in experiments of this type, and hence the developmental stages outlined, are constrained by the structure of the experiment, both in terms of its underlying 'logical' structure and the actual set up of materials and instructions used. A comprehensive discussion of this point is provided by Newell and Simon (1972), who introduce the concepts of 'problem space' and 'task environment'.

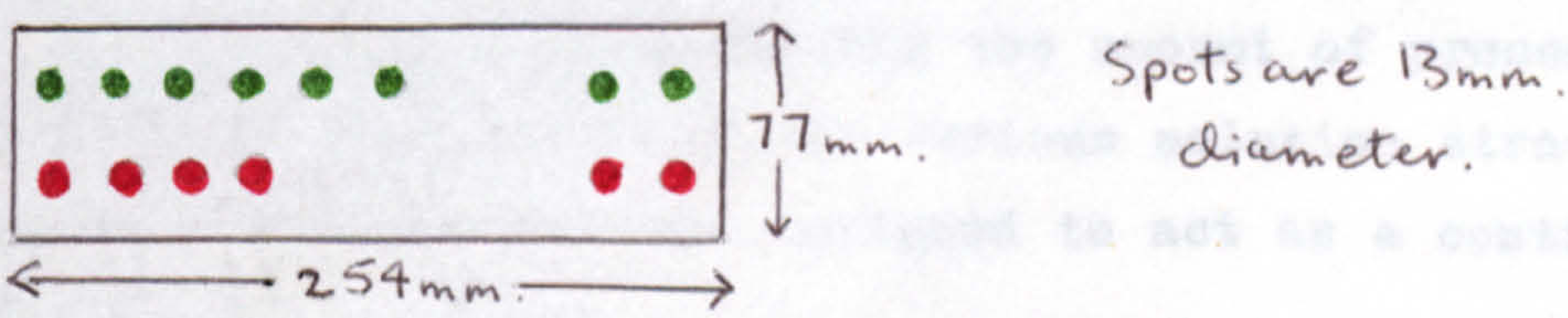
The context of Newell and Simon's discussion is an attempt to delineate concepts applicable to human information processing in order to prepare the ground for computer simulations of human problem solving. They point out that if there is such a thing as behaviour demanded by a situation, and if a subject exhibits it, then his behaviour tells more about the situation (task environment) than about him. All that is learnt about the subject is that he is in fact motivated toward the goal, and that he is capable of discovering and executing the behaviour called for by the situation. In a different situation he would behave differently.

The term 'task environment' is used by Newell and Simon to refer to an environment coupled with a goal, problem or task (for which the motivation of the subject is assumed). 'Problem space' is used to describe the subject's representation of the task environment, so that the construction of a problem space is an essential pre-requisite to any problem solving activity on the part of the subject. The importance of these concepts is well illustrated by an example drawn from Newell and Simon (1972):

> '1. To the extent that the behaviour is precisely what is called for by the situation, it will give us information about the task evirohment. By observing the behaviour of a grandmaster over a chessboard, we gain information about the structure of the problem space associated with the game of chess.
>
> 2. To the extent that the behaviour departs from perfect rationality, we gain information about the psychology of the subject, about the nature of the internal mechanisms that are limiting his performance.' (Op.cit., p. 55).

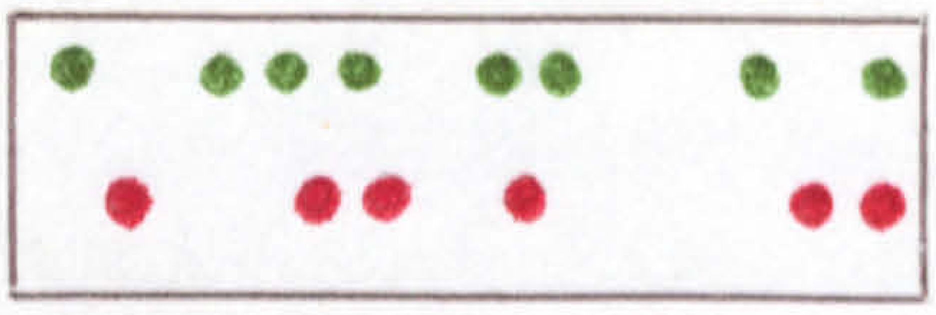
This distinction between demands of the task environment and psychology of the subject as aspects of the theory of problem solving is not, of course, meant to imply that psychologists should only concern themselves with the psychology of the subject. Newell and Simon insist that the two aspects are in fact like figure and ground, and that which is which depends on the momentary viewpoint. The importance of the distinction for present purposes is that it highlights two points which might otherwise be overlooked, namely that the task environment of the experiment constrains the set of strategies which subjects can usefully employ, and that the subjects must make an internal representation of the task environment in the form of a problem space before they can consider what solution strategy to employ at all.

This is not so much a criticism of the experiment as a caution against reaching premature conclusions as to the generality of results obtained in research of this type. One very glaring example of the way in which the experimental set-up might influence the solution strategies available to the subjects is apparent in the linear arrangement of the display cards used in the first beads experiment. The cards are used to act as a reminder to the subjects of the exact constitution of the collections of beads which have been placed in the two boxes, and the information contained on them is arranged in the form of two rows of beads in one-one correspondence as far as possible. For example, the experimental card for problem B4 is like this:

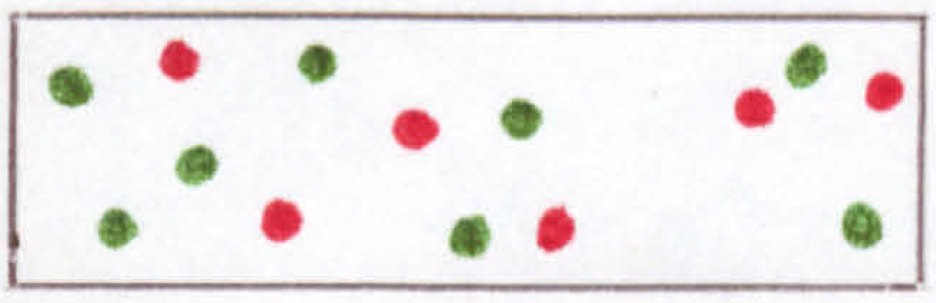


This layout was arrived at from considerations of the necessity of providing a clear and simple illustration of the situation. However, it is perhaps too striking in its demonstration of the fact that there are two extra green beads on the left-hand side, especially as it has been a common finding since Binet (1890) that one-one correspondence is a method of quantitative comparison available to most school children (see Bryant, 1971b; Brainerd, 1973b). This means that the stage 3 strategy of comparing the number of target-colour beads and non-target-colour beads in each collection may be facilitated or even induced by this method of representation.

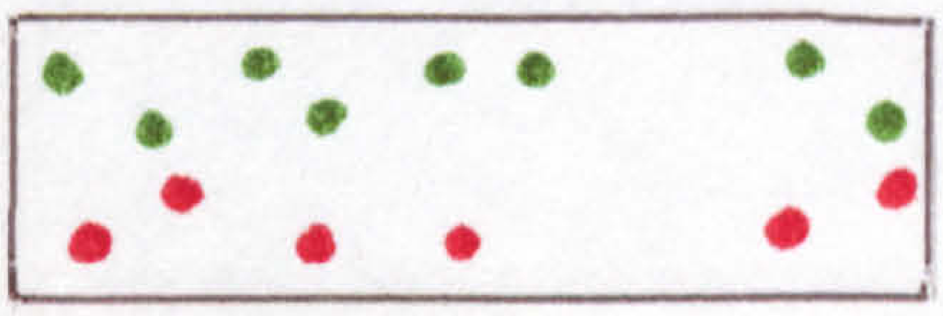
In order to test this possibility it was decided to carry out an experiment in which this 'seductive' layout would be avoided. There are various ways in which this can be achieved. The one-one correspondence of the beads in the two rows might be eliminated, but their linearity retained;



This is effective with the larger numbers involved in the left-hand side of problem B4, but is not so helpful with some of the problems involving smaller numbers of beads, where the linear display allows one-one pairing off to still be feasible without too much effort on the part of the subject. The representation could be modified to a 'mixture' type, in which the physical separation of the two collections of beads is maintained, but the red and green beads in each collection are scrambled:



This is undoubtedly more realistic, but it is no longer clear and simple. A compromise solution would be to destroy the linearity and one-one correspondence of the original displays, but maintain the physical separation of the red and green elements in each collection:



This seems to satisfy the need to remove the compelling aspects of the original cards without greatly increasing the amount of processing required for the successful application of the various solution strategies. An experiment using cards of this type was designed to act as a control on the result of first beads experiment.

<u>Procedure:</u> There were two differences in experimental procedure between the control experiment and the first beads experiment. The first difference was that a new set of experimental cards was produced to conform to the specification outlined above. In addition, the experimenter was allowed to ask either 'What do you mean by that?' or 'How many more are there?' to try and get subjects to clarify inexplicit choice reasons. The second question was not allowed in the original experiment on methodological grounds. However, as questions of this type are only rarely required, it was decided to allow it in this control experiment because it is a more precise way of disambiguating responses than the first question.

TABLE 4: <u>INTERPRETATIVE CATEGORISATIONS OF RESPONSES BY SUBJECTS IN THE CONTROL EXPERIMENT WHO HAD PREVIOUSLY BEEN INVOLVED IN THE FIRST BEADS EXPERIMENT, TOGETHER WITH THEIR RESPONSES TO THE FIRST BEADS EXPERIMENT</u>

	SUBJECT	PROBLEM C1 (B1)		PROBLEM C2 (B2)		PROBLEM C3 (B3)		PROBLEM C4 (B4)	
		GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
AGED 8 YRS. ON BOTH OCCASIONS.	SB43	2 (2)	2 (2)	2 (2)	2a (2a)	2 (2)	2 (2)	2 (2)	2 (2)
	SB44	2 (2)	2 (2)	2 (2)	2b (2b)	2 (3)	2 (3)	2 (2)	2? (2)
	SB45	2 (2)	2 (2)	2 (3)	2a (3)	3 (3)	3 (3)	3? (3)	2 (3)
AGED 8 WHEN TESTED FOR BEADS EXPERIMENT, 9 IN CONTROL EXPERIMENT.	SB37	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (2)	1 (1)
	SB46	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (2)
	SB47	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)
	SB48	3 (3)	3 (3)	3 (2)	3 (2b)	3 (3)	3 (3)	3 (2)	3 (3)
AGED 9 YRS. ON BOTH OCCASIONS.	SB49	3 (1)	3 (1)	3 (1)	3 (1)	? (1)	? (1)	3 (1)	? (1)
	SB53	2 (2)	2 (2)	2 (2)	2b (2b)	2 (2)	2 (2)	2 (2)	2 (2)
	SB54	2 (3)	2 (2)	2 (2)	2b (2b)	2 (3)	2 (3)	2 (2)	2 (?)
	SB57	3 (3)	3 (3)	3 (3)	3 (3)	3 (2)	3 (2)	3 (3)	3 (3)
	SB58	3 (3)	3 (3)	3 (3)	3 (3)	2 (2)	2 (2)	3 (3)	3 (3)

(FIGURES IN BRACKETS INDICATE RESPONSES FROM FIRST BEADS EXPERIMENT ITEMS).

Two groups of children took part in this experiment with the modified prentation of problems (which will be referred to as type (ii) problem presentation):

(a) Children who had taken part in the first beads experiment (involving type (i) problem presentation), and then took part in the control experiment (involving type (ii) problem presentation) as well:

Three children aged 8 years.

Nine children aged 9 years (four of these were 8 years old when they took part in the first beads experiment).

(b) Children who had not taken part in the first beads experiment (involving type (i) problem presentation):

Six children aged 8 years

Six children aged 9 years.

The reason for using eight and nine-year old subjects is that the control experiment is intended to establish whether the stage 3 type of response is facilitated or induced by the type (i) method of presentation, hence these ages, where the original experiment revealed the greatest mixture of stage 2 and stage 3 responses, are the most important for this purpose.

Results: The division of subjects in the control experiment into two groups, according to whether or not they had been involved in the first beads experiment, allows two independent comparisons of the results of the two experiments to be made. The results of the control experiment, scored according to the criteria outlined for the first beads experiment, can be found in Appendix D.

The first comparison to be made is the comparison of the results of the children who did both the first beads experiment and the control experiment. (The control experiment was carried out about 3 months after the first beads experiment). These are arranged in Table 4, and show little difference between the two sets of results, with the exception of one subject. Any differences which do exist take the form of improvements or regressions of one stage. There are six examples of regression on the control experiment, and six examples of improvements. In addition there is one example of an unclassifiable response to the first beads experiment which becomes a stage 2 response in the control experiment. This is out of a total of 88 responses produced by the 11 subjects conforming to this pattern, but there was one anomalous subject (SB49 in the first beads experiment) who gave all stage 1 responses in the first beads experiment, and then a mixture of stage 3 and unclassifiable responses in the control experiment. The extreme discrepancy between his strategy in each of these experiments is well illustrated by the following examples:

First beads experiment:

Problem B1 green.

Chooses 2G4R. 'Both the same size. The one I use looks bigger.'(Stage 1).

Problem B4 red.

Chooses 6G4R. 'I like the colours on the side.' (Stage 1).

Control experiment:

Problem C1 green.

Chooses 4G2R. 'It's got more greens than reds, the other is less.' (Stage 3).

Problem C4 red.

Chooses 2G2R. 'There's two of each. I say it's a medium chance. I don't want to get muddled up, so I'll get this. It's a draw, but I might win.' (Unclassifiable).

It would seem that this subject hadn't fully grasped what was required of him in the first beads experiment. This interpretation is supported by the fact that he was the only child over eight years old to give more than one stage 1 response in that experiment. The shift in his performance between the two experiments, but not within either experiment, also provides an interesting example of how little children will modify their solution strategy over a series of trials unless it leads to serious cognitive conflict.

Apart from this single subject the results of both experiments (in the case where each subject did both experiments) are closely comparable and support the conclusion that the stage 3 strategy is not dependent on the initial presentation format. However, the order in which the subjects did the experiments did not vary and it might be argued that the result is confounded with some sort of training effect, even though the two experiments were separated by a period of some three months.

This criticism is avoided in the second comparison, which is a comparison of results from two different sets of children who each did only one experiment. Six eight-year-olds and six nine-year-olds did the control experiment, and their performance can be compared with the performance of the eight and nine-year age groups in the first beads experiment (twelve of each) by comparing the proportion of the total number of responses of each stage produced by each age

TABLE 4(a): COMPARISON OF STRATEGIES USED BY EIGHT AND NINE YEAR OLD CHILDREN IN THE FIRST BEADS AND CONTROL EXPERIMENTS.

AGE GROUP.	% OF TYPE 1 ANSWERS.	% OF TYPE 2 ANSWERS	% OF TYPE 3 ANSWERS.	% OF UNCLASSIFIABLE ANSWERS.
8 YEAR OLDS. FIRST BEADS EXPERIMENT.	7%	64%	29%	0%
8 YEAR OLDS. CONTROL EXPERIMENT.	15%	65%	10%	10%
9 YEAR OLDS. FIRST BEADS EXPERIMENT.	11%	42%	39%	8%
9 YEAR OLDS. CONTROL EXPERIMENT.	0%	42%	48%	10%
COMBINED SCORES. 8+9 YEAR OLDS. FIRST BEADS EX'T.	9%	53%	34%	4%
COMBINED SCORES. 8+9 YEAR OLDS. CONTROL EXPERIMENT	7%	53%	29%	10%

(THIS TABLE IS COMPILED FROM RESULTS OBTAINED WHEN EACH SUBJECT ONLY DID ONE OF THE EXPERIMENTS) group. (This is preferable to randomly selecting six subjects from each of the age groups in the first beads experiment). This comparison is illustrated by table 4(a).

Table 4(a) shows a decrease in the proportion of stage 3 responses in the control experiment at age eight, and an increase at age nine, with an overall slight decrease. There is an appreciable increase in the proportion of unclassifiable responses, but the general spread of responses indicates that the three-stage model still provides a valid description of the results of the control experiment. This is the most important finding. Quantitative differences between experiments using different presentation styles are bound to occur, but as long as the differences are quantitative and not qualitative the conclusion can be drawn that the control experiment confirms the general model evolved from the first beads experiment. A statistical test of the 'significance' of the difference between the two treatments represented by table 4(a) has not been made, as there is clearly no significant difference.

In conclusion, then, the results of the first beads and control experiments are very similar, especially for the subjects who took part in both experiments. The subjects who only did the control experiment show a slight fall-off in performance as compared with subjects from the first beads experiment, especially at the lower of the two ages considered. This result is consistent with the view that the control experiment is slightly 'harder' than the first beads experiment, since the result in the case where subjects did both experiments may not show a drop in performance as a result of a few months age difference and increased familiarity with the experimental situation.

However, the overall pattern of results in both experimental conditions is similar and the claim that the presentation format of the first beads experiment induces the stage 3 strategy can be rejected. (An analogous result is reported by Wohlwill, 1968, who found that the spatial arrangement of experimental materials had no significant effect on class inclusion performance).

CHAPTER 3.

Theoretical considerations:

A digression concerning the word 'more':

Further pilot work was now undertaken to assess the feasibility of using the prod 'How many more?' to clarify ambiguous reasons. For example:

Collection A	Collection B
XXXX	XXX
YY	YY

When the target is a bead of type X, some children choose collection A and give as a reason 'Because there's more'. This might mean 'Because there's more X's than Y's or 'Becuase there's more X's here than in B' or it might mean something completely different. In the first beads experiment the question 'What do you mean by that?' was asked by the experimenter whenever such uncertain cases arose, in the hope that this might lead to a more detailed statement of his reason by the child. This tactic sometimes works, but sometimes it just leads to a repetition of the original response. The question 'How many more?' looks more promising because it is more specific and stage 2 reasoning leads to the answer 'One more' whereas stage 3 reasoning gives 'Two more'.

The difficulty with this kind of precise prodding is that it draws the children's attention to the quantitative aspects of the situation. This may lead to lack of flexibility in their reasons or, even worse, to the production of reasons which might otherwise have been overlooked. Because of this the question was not used in the first beads experiment, but it was allowed in the control experiment after the results of the first experiment had indicated how seldom it would have to be used. The reason for further pilot work in this area is that it was felt that a 'clinical' check on the value of this prod might be worthwhile.

One of the first things that became apparent was that some of the problems used in the first beads experiment are unsuitable for use with specific questions of the 'How many more' type. For example, problem B1:

Collection A	Collection B
XXXX	XX
YY	YYYY

is symmetrical in the sense that the number more X's in collection A than in collection B is the same as the number more X's than Y's in collection A.

The other three of the four main experimental problems do not suffer from this drawback, but additional problem B6 and the part of problem B7 where green is the target colour are also symmetrical. In problem B6, however, stage 2 and stage 3 lead to different choices, so that the consequences of the symmetry are unimportant, whereas in problem B1 and problem B7 green, both stage 2 and stage 3 lead to the same choice. Thus for future work problem B1 and problem B7 green will have to be replaced by better designed problems, since even if the question 'How many more?' is not used much information of this kind is given spontaneously by subjects and is being wasted.

Another reason for carrying out this pilot study was to provide a clinical assessment of the validity of using the subjects' verbal reports of their reasons for their choices as data in the first beads experiment and the control experiment. It is possible that in some cases children give answers which may suggest a different solution strategy to one they are really using.

For example, in problem B4:

Collection A	Collection B
XXXXXX	XX
YYYY	YY

when X represents the target colour, a subject might use the model 2 strategy to choose collection A, which has 6 X's as opposed to the 2 X's of collection B, and then justify the choice with a stage 3 reason, such as 'There's more X's than Y's'. In all problems where the model 2 and model 3 strategies lead to the same choice, this kind of juxtaposition of model 2 strategy and stage 3 reason, or model 3 strategy and stage 2 reason, can occur without any inconstency between the choice made and the reason given for the choice. The following problems allow model 2 and model 3 reasoning to lead to the same choice:

Problem B1, whether red or green is the target colour.
Problem B2, when green is the target colour.
Problem B4, when green is the target colour.
Problem B5, when green is the target oolour.
Problem B7, when green is the target colour.

However, in all the above cases except problem B1 and problem B7 green, it is possible to find out whether such juxtapositions occur by asking 'How many more?'

In order to assess the usefulness of this question a pilot study was carried out, using the non-linear presentation designed for the control experiment, with the experimenter asking the question 'How many more?' whenever possible. This is very time consuming, as the question is often inappropriate. The data gathered (from twelve children aged between five and eight years) revealed that the question is useful for clarifying ambiguous responses (although it may still be undesirable on other grounds), and that there doesn't seem to be any danger of artifacts in the first beads or control experiments caused by children saying one thing and meaning another. Children who give stage 2 or stage 3 responses seem to be very consistent in their use of them, and do not appear to juxtapose stage 2 reasons with stage 3 choices or stage 3 reasons with stage 2 choices. However, children who give stage 1 responses show much greater variability, and do seem to say things they don't mean. They give the impression of justifying their choices in a purely post hoc fashion, so that reasons which contradict the choice made, or reasons which might apply equally well to either collection, abound. In addition to providing clinical confirmation that verbal-response plus choice-made is a reliable source of data, the pilot work revealed an unexpected phenomenon which might be of importance concerning the use of the relational term 'more'.

It is well known that the way in which children use relational terms such as 'more' can be different to adult usage. (Donaldson and Balfour, 1968, Donaldson and Wales, 1970). However, the work cited involved younger children than were used in the experiments reported here, so that it was assumed that considerations of this kind would not affect them. Briefly, there are two points usually made about the way young children use the word 'more':

(a) They often fail to differentiate it from 'less'.

(b) They don't use it in a fully relational manner, but view the largest item of a group as being 'the more'.

Such effects tend to be highly dependent on what Campbell and Wales (1970) call the 'eliciting context'. By investigating children's comprehension and use of such expressions in as diverse a range of situations as possible, Wales (1971) is able to argue that what appear to be 'absolute' uses of relational terms in one context turn out to be comparative uses in different contexts. This means that point (b) must be treated carefully, and must not be taken as implying that children acting as if one part of a display is 'the more' are incapable of relational considerations.

The children involved in the first beads and control experiments who used the word 'more' all seemed to be using it in the adult fashion. That is, they said things like 'There's more here than there', 'This one has more than that one', and so on. But the pilot experiment revealed that for many children, although their linguistic usage is adult, the term may mean something slightly different to the adult meaning. What happens is that many children, when faced with two collections of beads, C andD, of which C has (C-D) more beads than D, claim that there are C more beads in C than there are in D. This appears to be quite common even with eight and nine-year-olds. Of course the 'eliciting context' in this case is quite complex, being the experimental task involving two collections of red and green beads. Further pilot work revealed,however, that the same effect can be observed in simpler contexts involving only two collections of beadsof a single colour.

From this it is possible to put forward the following ways in which children might use the word 'more' in situations involving two or more collections of objects:

(a) 'More' is applied to the larger collection, but is not differentiated from 'less'.

(b) 'More' is applied to the larger collection and differentiated from 'less', but used as if it was a property of the larger collection rather than a relation between the collections, e.g. 'The one there is more'.

(c) 'More' is used as a relation linguistically, e.g. 'There is more there than there', but when the difference must be quantified, that is when the question 'How many more are there?' is asked, it is not treated relationally and the answer given is the number of objects in the larger collection.

(d) 'More' is used as a relation linguistically and quantitatively.

It is tempting to regard this as a four-step developmental sequence, (although it can be argued that the distinction between (b) and (c) is overdrawn),but the evidence assembled here is too meagre to justify this. It would be interesting to know how use of the word 'less' develops after it is distinguished from 'more', as there are good reasons for supposing that the course of its development will not merely be mirror image of the development of 'more' (see Campbell and Wales, 1970 ; Wales, 1971). For present purposes, however this is unnecessary.

The reason for this digression is that the word 'more' features prominently in the reasons given by subjects in the experiments reported here. When used it seems to be used in the linguistically correct way, so that children giving reasons involving 'more' are beyond the steps (a) and (b) reported above (this is partly to be

Table 4 (b): Design of pilot experiment attempting to investigate possible relationships between children's understanding of the word 'more' and the solution strategies observed in the first beads experiment.

Pretest: Control experiment problems C2 and C4.
This is intended to roughly pinpoint the subject's level of ability.

Problems (including instructions):

A.

Collection A.	Collection B.	
YYY	YYYY	Y = green bead
XX	XXXX	X = red bead

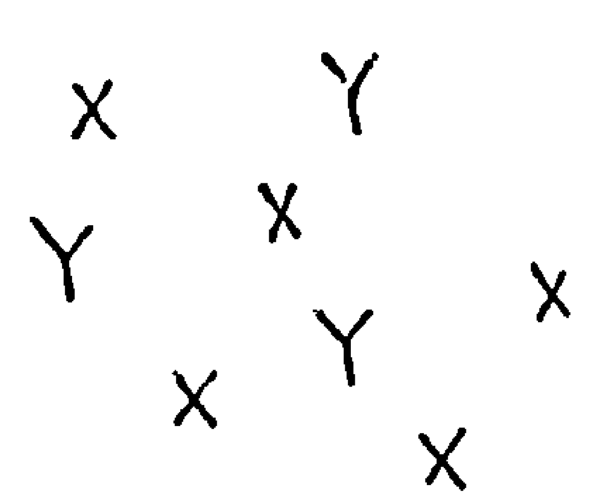
'We want to get a green bead.
Which box has more green beads than the other box?
Which box has more green beads than red beads?
Which box would you choose to get a green bead?'
The questions are then repeated, with red as the target colour.

B.

Collection A.	Collection B.
YYYY	YYY
XX	XXXX

'How many more green beads are there in this box?' (Point to Collection A).
'How many more green beads than red beads are there in this box?' (Point to Collection A).
'How many more green beads are there in this box than in this box?' (Point to Collection A, then Collection B).
'Which box would you choose to get a green bead?'
Repeat questions, with red as target colour.

C.



Y = green bead, placed in a box.
X = red bead, placed in same box.

'Are there more red beads or more green beads?
How many more?
Which colour will come out first if we tip the box?
Why?'

D. 'What is five take away three?'

expected, as the youngest children in the sample are aged 5 years). However, the possibility that they are not all beyond step (c) is tantalising. Could it be that steps (c) and (d) are related in a systematic way to the solution strategies children use in this kind of experiment?

An experiment to investigate this possibility was designed. It consisted of various problems to test both the strategies the subject would use to solve problems of the type used in the first beads experiment, and his comprehension of the word 'more' for linguistic and quantitative purposes. The problems used in the experiment are shown in table 4(b). After extensive pilot work (involving over thirty subjects aged between 5 and 8 years), and several revisions, the only conclusion to be drawn from this experiment was that 'everything is related to everything else'. No trends or systematic interrelationships were apparent, and the experiment was abandoned. This failure indicates the extreme difficulty involved in investigating language-thought interrelationships, and may well be attributable to inadequate design or oversimple aims.

Further development of the theoretical position

It is now necessary to consider in greater detail the theoretical position being developed in this thesis. The experiments reported involve decisions by children as to which of two chance events has the greater likelihood of a successful outcome. Models are then proposed to account for the decisions made and reasons given, whilst meeting the criteria of sufficiency (Gregg and Simon, 1967) and developmental tractability (Klahr and Wallace, 1970b). The reason for using the forced-choice situation is that it can provide an extremely sensitive source of data if it is carefully set up.

The problems used in the experiments concern what the Genevan school would call 'the child's conception of chance and probability'. Similar problems were in fact used by Piaget and Inhelder (1951). Unfortunately to talk in this way is misleading. It suggests that children would themselves classify the problems as concerning chance and probability, whereas this is really the experimenter's classification system. This sort of mistake leads ultimately to a 'cognitive phrenology' in which children have conceptions of number, space, time, causality, movement, speed, geometry, chance, probability, the world, morality, and anything else the investigator may care to name. In a trivial sense this is, of course, the case, but the point being made is that it is important that we understand the way in which the children in the experiments classify what is happening rather than imposing a ready-made classification system onto our theory. Examples of this tendency to read complex conceptions into performances on fixed (and often non-verbal) tasks can be found throughout the literature (e.g. Davies, 1965; Carlson, 1970).

There is nothing in these studies, nor in any of the experiments reported so far, to indicate that chance and probability correspond to the way the children involved classify such problems. Conversely, there isn't anything to suggest that this isn't the case either.

Ross (1966) seems initially to take this point, but then argues that there are certain types of behaviour which correspond to the use of probability concepts, and that certain tasks are more likely to lead to use of probability concepts than others. For example, in a task involving drawing balls out of a box and predicting which colour will be next when there are two possible colours and the number of balls of each colour in the box is known, Ross claims that alterations of predictions from trial to trial according to the results of previous trials will indicate the application of probability concepts. The exact meaning of 'probability concepts' in this sense is not clear and Ross's assertion that 'probability concept studies try to sample responses that show S's characteristic mode of responding in probability situations' (Ross, 1966, p. 917) doesn't help. It is parsimonious to suppose that until contrary evidence is produced such experiments only demonstrate the way children reason when faced with problems concerning what adults call chance and probability, not the child's conception of chance and probability, which might be something completely different.

What seems to happen in experiments of this kind is that the subject is faced with a complex problem of which he has at best only a limited understanding. He then has to use whatever means are available to him to produce an answer. If he can think of more than one way of obtaining an answer he will probably use the method he thinks most likely to produce whatever reaction he wants from the experimenter. The task thus becomes one of utilising problem-solving methods learnt in other situations, rather than drawing on knowledge of chance or probability (of course the appropriateness of methods learnt in other situations might be assessed by the child in terms of his conception of chance or probability, but this will not always be so). This is particularly the case with topics like probability, which are not normally encountered in school work, so that no situation-appropriate solution methods will have been taught to the child. (It is true that some schoolchildren are now encouraged to 'discover' probability theorems, but this was not the case in the schools selected for this study.)

This view is consistent with the information-processing analysis of Piagetian tasks undertaken by Klahr and Wallace (1970a), although it lacks their level of detail and precision. The idea that the subject will want to produce some kind of reaction from the experimenter is very important, and often overlooked. Hayes (1972) points out that the child's conception of theexperimenter can influence the result of an experiment in many ways. In fact it would probably be fair to say that the experimenter, being normally the only other person present, is the most important and the most problematic part of the child's task environment. The child must construct some sort of model of the experimenter for his problem space, and his behaviour throughout the experiment will be related to properties of this model.

Unfortunately there appears to be very little research on this interactive nature of experiments in developmental psychology, so that the present remarks will have to be confined to a few suppositions and anecdotes. Hayes (1972) starts from a position which seems to owe much to ethnomethodology, which is concerned with the way rules and meanings are negotiated in interaction by means of 'interpretive procedures' (Cicourel, 1973). He argues that in normal interaction the frequency of speech errors (and, one might add, allusions) means that we often have to 'fix up' what someone else says by interpreting what he meant rather than taking his words at face value, To do this the listener must have a good model of the speaker. Some quite common situations where the speaker and listener use the same word to refer to different concepts, such as the detection of irony, dishonesty, and misunderstanding, can put severe demands on the listener's ability to model the speaker. Similarly, the young child may be misled by an adult in the role of experimenter, because the adult is likely to ask questions not because he wants the indicated information, but because he wants to determine if the listener has that information. Unless this understood the child may well react by trying to patch up anomalies in what the experimenter said without being tactless enough to tell the experimenter he is doing so.

A case relating to the class inclusion problem where such an interpretation could be made, is culled by Hayes from Piaget (1952):

'Tap (5; 6)... Tap was then given two sets of beads in two boxes, each one contained 20 brown and 18 green all made of wood. "The little girl who has this box makes her necklace with the brown beads, and the girl who has the other box makes her necklace with the wooden beads in it. Which necklace will be longer?" - "The brown one because they are most." - "And what colour will the necklace of wooden beads be?" - "Only green". (Piaget 1952, p. 169)

Hayes comments '.....it strikes me that these very same results also cast doubt on Piaget's own interpretation of the class inclusion data - namely that the child cannot simultaneously conceive of a class and a subclass of that class. In this new version of the problem, the child is not required to conceive class and subclass simultaneously - yet he still fails. These results, however, are consistent with the interpretation that the child treats the experimenter's question as anomalous and fixes it up.' (Hayes, 1972, p.180)

The same interpretation might also be applied to an observation by Inhelder (1972):

> 'When faced with a bunch of flowers containing a great may roses and a few tulips, and asked if there are more roses or more flowers, the child replies that there are more roses. If he is then asked "more than what?" he often answers "than tulips." (Op. cit., p.111).

Another example, where an inappropriate model of the experimenter clearly leads to faulty interaction,is given by Wales (1971),

> 'A boy of four is asked how many legs a horse has, and he refuses to answer. Margaret Donaldson, knowing that the boy knew the correct answer asked him afterwards why he had not given it - "If that big man didn't know, then I wasn't going to tell him!" (Op. cit., p. 73).

In the present research an example was found (SR26 in a later experiment) where the child either failed to construct an adequate model of what the experimenter wanted, or else overruled his model of the experimenter's wishes because of his own interests,

E: 'Why did you choose that one?'
SR26: 'Don't know why. Why do you keep asking me why?'
E: 'I'm interested'.
SR26: 'Well I'm not!'

In another experiment the same child adopted a different deviant position,

E: 'Why did you choose that one?'
SMR26: 'I want to..... Hee Ha! He writes down everything I say!'

Since most psychologists have failed to concern themselves with such problems these anecdotes can act as little more than a caution against underestimating the social nature of experiments.

There appear to be two main things the subject must do in any experiment of the type under discussion:

(a) He must encode the information he is going to have to deal with. In Newell and Simon's (1972) terms,he must construct a problem-space when faced with the task-environment. This problem-space will include a model of the experimenter as well as the experimental task.

(b) He must select a strategy of some kind in order to derive an answer from the encoded information.

Obviously there will be many constraints on the subject's ability to do these things, not the least of which will be the amount of information he can keep accessible at any time. The possible role of different types of memory store in the development of problem-solving abilities is given a clear and simple exposition by Simon(1972).

In more general terms (a) might be called representation and (b) problem-solving strategy. Since representation occupies such a central position in problem-solving it is surprising how little attention it has received in comparison with problem-solving strategy. Miller (1956 a, 1956 b) has indicated how coding can affect memory and the information handling capacity of the brain, and the same topic is given a more extended treatment by Neisser (1967). In Simon and Siklóssy, (1972) a number of possible representations of task environments are discussed, but these are mostly unsuitable for present purposes as they tend to be excessively formalised because of their direct link with attempts to model human behaviour using computers.

A psychological theory which is explicitly concerned with representation, and which is sufficiently flexible to be worth considering is that of Kelly (1955). Kelly holds that people represent things to themselves in terms of a system of dichotomous constructs, each of which is applicable within a certain range of convenience, and that this representation is used as a basis for predicting events and deciding what to do. This approach is useful because it provides a way of thinking about what the child's model of the experimenter might be like. Unfortunately, the theory is only applicable to representations corresponding to evaluations made by the subject, which is why it appears to break down the cognitive-affective distinction. It is difficult to imagine how the more 'objective' aspects of situations, such as the fact that there are, say, red and green beads jumbled up in a box, might be represented in terms of constructs.

Bruner (1966) provides the basis for a more general theory of representation. He maintains that there are three main modes of representation, which he calls enactive, **ikonic**, and symbolic. Briefly, enactive representation is a way of representing things by the actions that one can perform on them, **ikonic** representation involves an image or spatial scheme that is relatively independent of action, and symbolic representation involves some kind of arbitrary or conventional symbol referring to the thing signified. It can be seen that constructs constitute one form of symbolic representation, since although Kelly maintains explicitly that all constructs need not have names, the relation of construct to whatever is construed can be seen as similar to the relation of symbol to significate. However, Bruner suggests that any symbolic activity is logically and empirically unthinkable without the properties of categoriality, **hierarchy**, predication, causation, and modification, but this seems to rest more on an analogy with the analysis of natural language systems by Hockett (1960) and Chomsky (1965) than any more pertinent considerations.

Bruner claims that the modes of representation develop in the order in which they have been presented, and that much of the course of cognitive development can be understood in terms of their development, interaction, and integration. He argues that when systems of representation come into conflict or contradiction the child is likely to make sharp revisions in his way of solving problems, for instance when there is a discrepancy between 'appearances' (ikonic) and 'reality' (symbolic).

In the problems under consideration at present one of the first steps in the formation of an internal representation of the actual task must be to separate the two collections of beads and the red and green beads in each collection. This might well be viewed as a function of perception, without needing to claim that perception is active, passive, continuous, categorical, innate, learnt, or whatever. In other words the only claim made is that all children of the ages used in the experiments are capable of doing this and will do it, unless they are colourblind. A choice of enactive, ikonic, or symbolic mode of representation must also be made, possibly before this 'perceptual processing', possibly after. It follows from Bruner's arguement that the younger children in the experiment (ages 5 and 6) are likely to use ikonic or symbolic representations, whereas the older children will probably rely on the symbolic mode. An initial ikonic image of collections and colours can easily be translated into symbolic form for this purpose by some such expedient as labelling the collections 'left'and'right' and the beads 'red' and 'green'. This however, is only one of a variety of possible symbolic representations.

Another possibility, used by Klahr and Wallace (1972), is to represent the objects in the set of materials by lists of values, which in this case would mean a list of objects in terms of two values (colour and collection-membership) plus a list serving as a record of the values appearing in the set. The list of objects might then be transformed into four lists representing the interaction of the two values by assigning each object on the original list to one of the new lists (note that this is only possible when the resulting four lists are mutually exclusive in this way, and no object will appear on two of the new lists. In problems which would require assignment of the same object to two or more lists, such as class-inclusion problems, a more complicated way of constructing the new lists would be necessary, e.g. Wallace, 1973).

A very powerful improvement can be made to the list representation if the introduction of quantitative symbols which can be compared in magnitude by appropriate operations is permitted. This is a point of general importance as many kinds of representation are amenable to quantification, and the quantification process itself generates new symbols. Klahr and Wallace (1973) introduce the idea of quantification operators to describe this process,

> 'A quantification operator is an organized collection of elementary processes that takes as input the stimulus to be quantified (e.g. a collection of blocks) as well as specified constraints (e.g. red only) and produces as output a quantitative symbol. Quantitative symbols are labeled internal representations (e.g. "two," "long," "tiny") that can be used in quantitative comparisons. Given two such symbols, the organism can determine their relative magnitudes...' (Op.cit., p.303).

The existence of three quantification operators, subitizing, counting and estimation, is postulated by Klahr and Wallace. The evidence for these operators is drawn from analyses of the reaction times and errors of adult subjects in tasks requiring them to report the number of items in a display. These experiments show a monotomic increase in reaction time with increasing number of elements in the display in the range btween one and about thirty items. The slope of this increase is approximately 40 milliseconds per item for the range between one and five items, and approximately 300 milliseconds for the rest of the range (Klahr, 1973; but see also Woodworth and Schlosberg, 1954). The subitizing operator, for which the symbol Q_s is used, is postulated to account for the 40 millisecond slope, the counting operator, Q_c accounts for the 300 millisecond slope, and the estimation operator, Q_e is introduced to account for what happens outside the one item to thirty items range where there appears to be zero slope.

In adults Q_s, Q_c, and Q_e are fully developed, so that Q_s is used for quantifications involving a small number of items and Q_e is used for quantifications involving a large number of items or continuous quantity, leaving Q_c for the intermediate situations. However, the development of the three quantification operators seems to be very complex. There is some agreement that Q_s and Q_c develop concurrently whereas Q_e is a later achievement. The exact relationship between the development of Q_s and Q_c is more problematic. It is a general principle of the developmental theory advanced by Klahr and Wallace (1973) that if the child is viewed as a developing information processing system then that system will constantly search for any consistent sequences which will enable it to eliminate redundant processing. This means that the concurrent development of two reliable quantification operators will lead to a good deal of interaction between these two developments in the form of mapping of the properties of one system onto the other.

Gelman (1972a, 1972b) has claimed that counting precedes subitizing. She defines subitizing by the absence of overt counting, and has shown that by this criterion children aged 4 to 6 years are more likely to count than subitize the younger they are (with displays containing 3 to 6 items). Klahr and Wallace (1973), dispute this, claiming that Gelman's criterion underestimates subitizing in cases where both Q_s and Q_c are used. Fortunately the details of this disagreement are not pertinent to the present discussion as long as it is accepted that by age 5 or 6 both Q_s and Q_c are sufficiently developed and co-ordinated to lead to consistent results within the range where they overlap, whereas Q_e is not.

The existence of some quantification operator other than Q_s and Q_c is clear, because quantitative symbols can be produced in situations involving great numbers, or limited exposure duration, or continuous quantity, where neither Q_s nor Q_c could function. Wallace (1972b) goes so far as to claim that Q_e is in operation along with Q_s before Q_c. If this is the case then while Wallace's child is detecting consistent sequences between Q_s and Q_c, he would seem to be failing to detect consistent sequences between Q_s and Q_e, or between Q_c and Q_e. This failure is attributed by Klahr and Wallace (1973) to the fact that the wider scope of Q_e means that it can be applied to a number of different dimensions in most situations. Initial attempts to compare the results of Q_s and Q_c with quantitative symbols generated by applying Q_e to only one dimension of a situation are thus bound to lead to apparent inconsistencies.

The second general principle of Klahr and Wallace's developmental system is that if the system fails to discover consistent sequences in a particular context it will widen the basis of its search. This means that it will begin to apply Q_e to more than one dimension of a given situation, and in this way co-ordinations can be discovered which will allow reconciliation of the discrepancy between Q_e and the other operators, Q_s and Q_c, by a further mapping of properties from system to system.

This outline of Klahr and Wallace's ideas has been of necessity brief and vague, and much greater detail can be found in the references cited. The notion of quantification operators has been introduced into the present discussion because it seems to clarify certain aspects of the results of the experiments being considered. References to explicit or implicit quantitative symbols and comparisons of the magnitudes of sets of items represented by quantitative symbols were very common in both the first beads and control experiments. In fact all reasons corresponding to models 2, 3, and 4, and some stage 1 reasons involved reference to some sort of quantification processes. This is of course circular, because one of the defining properties of models 2, 3, and 4 is that quantitative comparisons are made, but it does draw attention to the importance of quantification processes in the children's performance.

At this point the observant reader will have noticed that the discussion has drifted a long way from Bruner's (1966) position, since the three quantification operators, as described by Klahr and Wallace (1973), seem to violate his three modes of representation. The quantitative symbols generated by counting (numbers) are clearly symbolic and culturally transmitted. Subitizing is often regarded as largely perceptually based, but Wallace (1972b) argues that consideration of the information processing required for a sufficient Q_s indicates that this is a misconception. The quantitative symbols generated by Q_s are at first personal but become mapped onto the culturally transmitted number system as links between Q_s and Q_c are established. Estimation might seem to be purely ikonic, but Klahr and Wallace (1973) maintain that like any measurement it must involve repeated application of a 'standard' unit. These standard untis will be based upon the experience of the particular system, and will thus be personal sybols. Some people may estimate lengths in terms of football fields, others in terms of cars, and so on. Wallace (1972b) calls these size analogue symbols. Repeated application of Q_e to different dimensions of a situation can result in its representation by a variety of size analogue symbols.

It seems that much of this apparent contradiction may be attributable to different uses of the terms 'representation' and 'symbol'. The way these words are used at present is very much a matter of individual style, but clear awareness of the various possibilities is shown by Werner and Kaplan (1963),

> 'The term "representation" is a rather common word used in various areas of discourse. In the area of cognition, the term is sometimes used to designate the relation between an abstract concept and a concrete example: one may thus be said to "represent" the concept tree by a percept or image of a particular tree. This "exemplificatory representation" differs clearly from symbolic representation: the concrete object, tree, is a "substantialization" of the connotations of the concept but does not symbolize or depict the concept; it would be preferable to speak here of "reification" instead of "representation" of a concept. In contrast, the word "tree" truly signifies or symbolizes the concept tree: the dynamic structuring of the word by the speaker (and the hearer) is taken within the linguistic medium to correspond to the significate (connotational structure of the referent). In order for such correspondence between vehicles and significates to be attained, the sybolic vehicles - e.g. the word-forms - have to be constructed systematically; language has become the medium of representation par excellence precisely because its vehicular forms, from the phonemic sound elements to the most complex syntactic structures, are built on systematic principles, making it possible to reveal, within the linguistic domain and by genuinely linguistic devices, the connotational structure of the referent'. (Op.cit., pp. 15-16, underlined word italicised in original).

The distinctions drawn in this extract are most useful. The kind of representation referred to by Werner and Kaplan as symbolic representation seems to bear a relationship to Bruner's symbolic mode in that the symbols refer to concepts and are organised into a language system. At the same time they make clear that the kind of representation referred to by Bruner as ikonic is not merely a different mode to symbolic representation, it also serves a different purpose. This is touched on by Bruner (1966) when he claims that representation can be understood in two senses, in terms of the medium employed and its objective, but he plays down the qualitative differences between his modes of representation in order to emphasize the idea that partial translation between modes is possible.

It is notable that Werner and Kaplan use the term symbol only with respect to concepts. They stress the wholistic aspects of concept symbolisation and its systemic properties. The type of symbol involved in Klahr and Wallace's (1973) notion of quantitative symbols does not seem to fit this type, and emphasizes the abstractive and analytic side of things.

Three steps will be taken here to clear away some of these difficulties. Firstly, the term representation will be used in the 'loose' sense advocated by Bruner, although the importance of the qualitative differences between different types of representation highlighted by Werner and Kaplan must not be ignored. Secondly, it will be conceded that symbolic representation refers to a variety of phenomena, ranging from concept systems to the products of operations, with the symbols transmitted culturally or defined from personal experience. Thirdly, it is suggested that many problems can be avoided if levels of representation are considered. Klahr and Wallace (1973) maintain that a quantification operator takes as input the stimulus to be quantified and certain specified constraints producing as output a quantitative symbol. If it is remembered that the input must take the form of an internal representation of the stimulus to be quantified together with the necessary constraints, the apparent disagreement with Bruner's (1966) theory can be avoided (since this can be seen as applicable to the initial representation), and the special nature of quantification processes as operations on inputs is emphasized. This formulation also draws attention to the fact that whilst quantification may be an important way of transforming inputs into manipulable representations, it cannot be applied without prior representation of at least the constraints on its application, i.e. quantification processes can only be considered in the context of a more general representation. An easy way of digesting this is to think in terms of Bruner's translation analogy and consider that quantification processes provide a way of recoding representations into a special symbolic form.

Returning to the experiments under consideration, it can be seen that the problems involve sets of items within the counting range of most subjects in the study, and sometimes within their subitizing range, sometimes not. The main experimental problems have sets of beads varying in numerosity between two and six, and two of the additional problems involve a set of eight beads. The youngest children in the experiments are five years old, by which age subitizing up to three or four items might reasonably be expected, together with counting in the complete range used (although this need not imply precise co-ordination of subitizing and counting). With the subjects aged seven years and above it is safe to assume subitizing ability in the range up to four or five items and accurate co-ordination of subitizing and counting (by making a pessimistic interpretation of the evidence presented by Klahr and Wallace, 1973, and Klahr, 1973).

This means that all of the subjects aged over seven years should have available to them a reliable method of quantification for all of the problems used in the experiment, and subjects below this age will often be in the same position. The reliability of Q_s and Q_c in this way (irrespective of their hypothesised relationship to each other) means that the fact that items in some of the parts of the experimental problems can be subitized or counted, whereas others can only be counted, should not affect the experimental results at the level of analysis employed. The fact that some of the younger children may not be in possession of some reliable method of quantification covering the whole range from one to six items (five and six year olds were not normally given the additional problems) is more important.

However, it should not be assumed that the absence of a reliable quantification operator compels the child to adopt the stage 1 type of solution strategy, since there is nothing in models 2, 3, or 4 which stipulates that reliable quantifications have to be made. Estimation could be used in any of the problems, but is likely to be ignored when more reliable methods are available. In certain of the problems, however, many subjects would be able to use Q_e reliably. For example, in problem B4 when green is the target colour, there are six green beads in one box and two green beads in the other box; with a discrepancy of this size most children would be able to use Q_c to deduce which box contained more green beads and thus avoid having to count or subitize. This is particularly likely to happen with the original 'linear' problem displays, where length of line is the only dimension which needs to be considered in order to generate a reliable quantity estimate. In the control experiment problems a compensation between length and density would be necessary, whilst the irrelevant non-linearity of the displays would have to be ignored.

Investigations of problem-solving strategies are more common in the research literature than attempts ot investigate forms of representation, possibly because strategies are more amenable to current experimental techniques. The first beads and control experiments are in line with this trend in that they were designed and analysed in order to investigate the applicability of various models of problem-solving strategies. Thus they contain the implicit assumption that the distinction between representation and strategy is valid, and that any interaction between the two will not be of a kind which might contaminate the type of analysis employed.

The present theoretical considerations support the view that the separation of representation and strategy is valid for analytic purposes, but indicate that the distinction is by no means as sharp as was initially supposed. There appears to be a close tie between the strategies distinguished in the beads experiments and quantification processes. The three modelled strategies, models 2, 3, and 4, all involve various quantifications and the exact nature of these quantifications is dictated by the requirements of each strategy. As an example of this, the model 2 strategy involves comparison of the amount of target-colour beads in each of the two collections, and selection of the collection with the larger amount. This means that some kind of quantification process is necessary to determine the amount of target beads in each collection, but the amount of non-target beads in each collection need not be quantified. If representation is considered as completely separate from strategy, the representation would have to be in a form suitable for the strategy chosen to be applied to it, so that the amounts of both target and non-target beads in each collection would have to be quantified (as it cannot be known in advance what the particular requirements of the strategy will be). This may in fact happen, but it seems more likely that the necessary quantifications will be dictated by direct strategic considerations, so that the amount of non-target beads in each collection need not be quantified when using the model 2 strategy. This would be consistent with the view expressed earlier that quantification processes provide a way of recoding representations into the more specialised form of quantitative symbols. At present, however, it is not possible to bring empirical evidence to bear on this point.

The overall model of what the subject does in the experimental situation is now as follows:

(a) Initial representation of the situation and its requirements.
(b) Selection or assembly of an appropriate strategy for coping with the requirements of the situation.
(c) Any recoding of initial representations necessary for the application of the chosen strategy.

(d) Execution of the chosen strategy.

Within this general framework an attempt has been made to elucidate the strategies used by children in a particular type of situation, and to introduce the idea of quantification operators as a way of viewing the quantifications carried out in the execution of these strategies.

CHAPTER 4

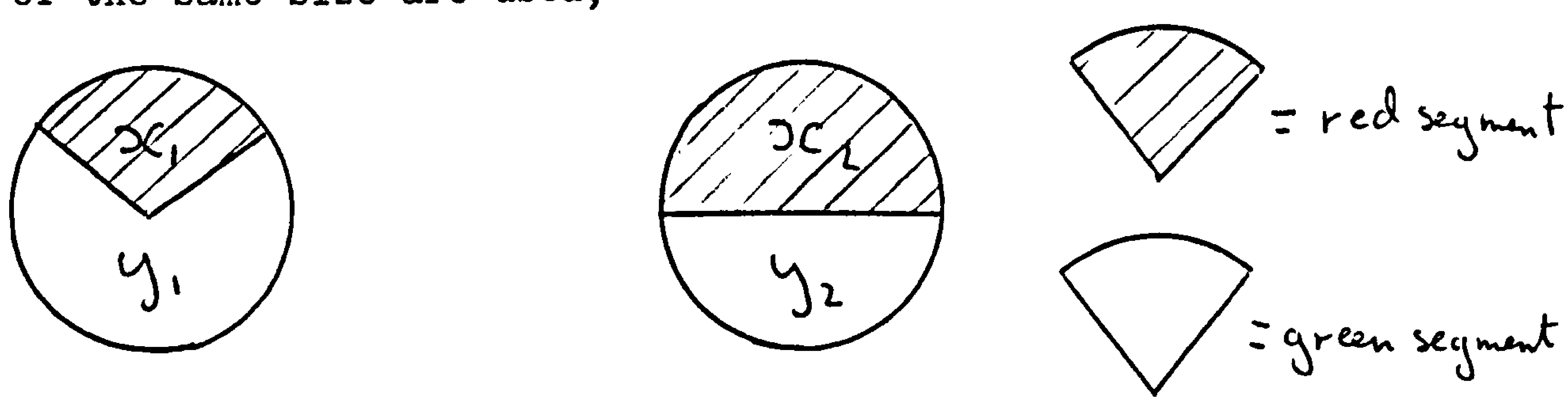
The Roulette Experiment

First thoughts and pilot work for the roulette experiment.

At this point it would be interesting to try out these ideas by seeing how well they can be applied to different experiments in related fields of enquiry. Unfortunately, this isn't possible with most of the extant work as the details which would be required in order to make a precise application aren't always available. Instead, an experiment will be designed to test out the ideas developed so far in a different context. It is quite possible that the strategies outlined from the results of the beads experiments may be situation specific. This is particularly likely with the model 3 strategy, which was not reported by Piaget and Inhelder (1951), and was found to be infrequent in the experiments carried out by Fischbein et. al (1970). The new experiment can be used to assess both the generality of the strategies identified and the usefulness of the theoretical position which has been developed.

What is needed is a type of problem which appears different to the problems already used, but is conceptually similar (to adults). This could be achieved most simply by changing the materials used, for example dice or roulette wheels might replace beads in boxes. Dice, however, are not very satisfactory as some faces will always be hidden from the subject doing the experiment, and even if a two-dimensional representation of what is on the faces is set up (as in the 'beads' experiment), it will be tricky establishing that this is really like what is on the dice. Roulette wheels with different proportions of red and green marked on their faces are more suitable in this respect, and also allow the use of continuous instead of discontinuous quantities.

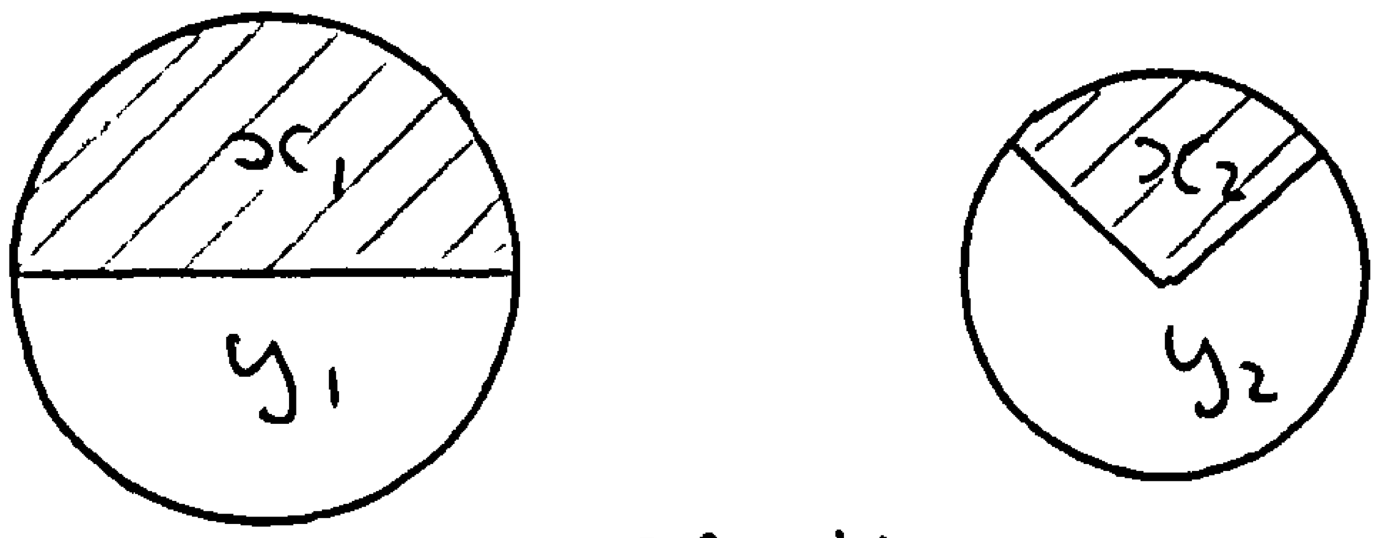
In order to arrive at a similar 'logical' set-up to the beads experiments it is necessary to use roulette wheels of different sizes. (The term 'roulette wheel' is used to refer to a pointer spinning on a circular face. This is not quite the same as the casino arrangement). If wheels of the same size are used,



x_1, x_2, are quantitative symbols representing the sizes of the red segments.
y_1, y_2, are quantitative symbols representing the sizes of the green segments.

It can be seen that $x_1 + y_1 = x_2 + y_2$

This means that there is an inverse relationship between the sizes of the red and green segments on each wheel, which causes the model 2 strategy of choosing the wheel with more of the target colour to always lead to the choice which is mathematically most likely to succeed (as long as a valid and reliable way of quantifying the sizes of the segments is discovered). As long as this happens the model 3 strategy is spurious. This doesn't necessarily mean that children will not use it, but that with wheels of the same sizes the model 2, 3, and 4 strategies will always lead to the same choices. In order to avoid this it is necessary to have wheels of different sizes, in which case,



$$x_1 + y_1 = x_2 + y_2$$
$$\text{or } x_1 + y_1 \neq x_2 + y_2$$

depending on the way in which x_1, x_2, y_1, y_2, are obtained.

e.g., if x_1, x_2, y_1, y_2, are fractions then $x_1 + y_1 = x_2 + y_2$

If x_1, x_2, y_1, y_2, are size estimates then $x_1 + y_1 \neq x_2 + y_2$

The possibility of using continuous quantities is very interesting in view of the theoretical position developed so far. It has been maintained that the use of small discontinuous quantities in the beads experiments led to quantification mainly by subitizing and counting, Q_s and Q_c. With continuous quantities Q_s and Q_c are inappropriate, and estimation (Q_e) will have to be used. However, Q_e is unreliable by comparison with Q_s and Q_c, especially at the younger ages used in the beads experiments, which leads to the prediction that the use of continuous quantities will lead to a result which will be 'retarded' with respect to the results of the beads experiments.

As an aid to designing an adequate experiment, a short pilot study was carried out to see if the use of roulette wheels of different sizes was feasible, and what methods children would use to encode information of this kind. The materials used were cards with wheels of different sizes marked on them. Each wheel was divided into a red and a green segment by means of shading lines, and a pointer was placed in the centre of the wheel which, after being spun, would finish pointing to the red or green shaded segment. The wheels used were either 165 mm. in diameter or 280 mm. in diameter, the shading lines were 13 mm. apart, the fractions of green used were $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$ (and hence $\frac{3}{4}$, $\frac{2}{3}$, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ red). Subjects (seventeen children aged from 7 to 11 years) were offered a choice between a gamble with one of the large wheels (chosen by the experimenter)or a gamble with one of the small wheels to get a specified colour.

The results of this pilot experiment will not be reported in detail as they conform to the pattern found in the more elaborate experiment about to be described. However, it is worth drawing attention to the following methods of quantification which were referred to by subjects in the pilot experiment (some of the same methods can also be found in the sample protocols of subjects in the roulette experiment proper):

(i) Counting the number of shading lines on a particular segment or the whole circle.

e.g. SRP14, $\frac{1}{3}$ RS v $\frac{1}{3}$ RL, green target.

(Pilot experiment subject number 14,offered a choice of small wheel with a red segment covering a third of its surface or a large wheel with a red segment covering a third of its surface,to get green).

Chooses $\frac{1}{3}$ RL. 'It's got more lines than the other one. Fifteen onto nine'.

(ii) Estimating the area occupied by a particular segment (its' size) or by the whole circle.

e.g. SRP 3, $\frac{1}{2}$ RS v $\frac{1}{4}$ RL, green target.

Chooses $\frac{1}{4}$ RL. 'It's bigger than that one. The circle is big.'

SRP8, $\frac{2}{3}$ RS v $\frac{1}{3}$ RL, red target.

Chooses $\frac{2}{3}$ RS. 'The other hasn't got as much red space as this one.'

(iii) Estimating the width of a segment.

e.g. SRP13, $\frac{3}{4}$ RS v $\frac{2}{3}$ RL, green target.

Chooses $\frac{2}{3}$ RL. 'More green than on the other one 'cos it's a wider bit'.

(iv) Estimating the size of the angle a particular segment substends at the centre of the wheel.

e.g. SRP17, $\frac{2}{3}$ RS v $\frac{1}{3}$ RL, red target.

Chooses $\frac{2}{3}$ RS. 'It's got more spread than the other one. Goes round further'.

(v) Recognising a certain shape of segment (or range of shapes)as corresponding to a particular fraction.

e.g. SRP11, $\frac{1}{2}$RS v $\frac{2}{3}$ RL, red target.

Chooses $\frac{2}{3}$ RL. 'There's three-quarters of red there and only half of red here'.

(vi) A more elaborate use of fractions in which the necessity of the two segments on each wheel adding up to one unit is realised, accompanied by attempts to calculate rather than recognise fractions.

e.g. SRP12, $\frac{2}{3}$ RS v $\frac{3}{4}$ RL, red target.
Chooses $\frac{3}{4}$ RL. 'It's only got a quarter green, so three-quarter is red. That one is thirds'.

In terms of Klahr and Wallace's (1973) proposed quantification operators, there is no sign of Q_s in the above list. Counting the number of shading lines corresponds to Q_c applied to an inappropriate feature (from the adult point of view). The next three methods in the list, (ii) - (iv), are ways of applying Q_e. Method (ii) (size estimation) is the most primitive as it involves more than one dimension, and Klahr and Wallace (1973) have argued cogently the need for viewing the development of Q_e as one of increasing analytic separation of the various possible dimensions to which it can be applied. The early appearance of method (ii) is confirmed by the fact that it is often applied to the size of a whole wheel rather than a particular segment, and this corresponds to the stage 1 outlined in the beads experiments. The more refined versions of Q_e were not used in this manner, but Q_c (as applied to the number of shading lines) sometimes was.

Some subjects in the pilot experiment used what at first sight appears to be a different method of quantification to those appearing in the above list, but may be a special case of the size estimation method. This involved a failure to discriminate anything except which segment of a wheel was larger than the other (i.e. which segment occupied more than half the circle). When faced with two wheels, one having $\frac{1}{4}$ red and $\frac{3}{4}$ green marked on, the other $\frac{1}{3}$ red and $\frac{2}{3}$ green, a child using this method would give the impression that as far as he was concerned there was more green than red on each wheel and nothing more to be said. It is difficult to be sure that this is really what is happening, and that the child is not just confused or holding back, but it can be seen as an early form of Q_e in which what Wallace (1972b) calls size analogue symbols are not well developed, and only relational attributes are coded. (Evidence is presented by Bryant, 1974, suggesting that young children do code relations in this way).

Estimates of the widths of segments, method (iii), are an improvement over size estimates in that only one dimension of the situation is being estimated. Method (iv), estimating the size of the angle a segment substends at the centre of the wheel, is similar in this respect but represents an advance by adult criteria because this is an aspect of the display which is relevant to a satisfactory solution of the problems, whereas sizes and widths of segments are irrelevant. A method of quantification which seems to be a variation of method (iv), and so was not included in the list, can sometimes be observed in cases where the two different-sized wheels have red (and green) segments occupying the same fractions of their surfaces. This can elicit comments like 'They look the same' or 'looks like the same on each'. Such decisions could only be based on comparisons of angles subtended at the centre, or segment shapes, which amounts to the same thing, (the only difference between the shapes of segments of a circle is in terms of the angle subtended at the centre).

The use of methods of quantification involving fractions might be thought to be very different to estimation. Estimation is an imprecise tool, and the development of Q_e seems to consist to a large extent of attempts to improve its reliability by specifying precisely the constraints under whibh it will be used. In spite of this it remains essentially imprecise. Fractions, on the other hand, are a very precise quantitative symbol implying some depth of relationships. However, method (v), the recognition of fractions, is arguably more like estimation than true fractional quantification. This is because the fraction is only determined by what it looks like, so that misquantification is possible if the child only knows one or two fractions. Thus, in the pilot study, examples are found of children who only seem to know of the fractions $\frac{1}{2}$ and $\frac{1}{3}$ claiming that segments representing both $\frac{1}{3}$ and $\frac{1}{4}$ of the whole wheel are $\frac{1}{3}$, and hence failing to distinguish them. When used in this way a fraction is no more than a culturally transmitted size analogue symbol, and this interpretation is supported by statements of the type 'that's more than a half, but less than a quarter', which can be found in the results of the pilot experiment.

True fractional quantification, unlike recognition of fractions, involves a certain amount of calculation. Because of this it has a different logical status to the other methods of quantification which have been commented on. These other methods can all be seen as forms of Q_c or Q_e, but fractional quantification does not correspond to any quantification operator, being instead something that can be done with a set of data which has already been quantified. With continuous quantities there is thus no precise way of arriving at 'true' fractions (unless one hands out protractors, or some other sort of measure is available) since the only initial quantification possible is in terms of Q_e. This means that methods (v) and (vi), recognition of fractions and appreciation of the real nature of fractions, are sometimes indistinguishable, particularly as so many of the relationships expressed in a fraction remain implicit. The importance of this point will become clearer later, when the results of the experiment being led up to are discussed. At present it will suffice to say that although the distinction between methods (v) and (vi) is sometimes blurred it is still valid, as is illustrated by cases where children merely recognising fractions can claim that a wheel has $\frac{1}{2}$ red and $\frac{1}{3}$ green, or $\frac{1}{3}$ red and $\frac{3}{4}$ green, and so on. The children who realise that a fraction expresses a relation of part to whole are unlikely to say such things.

In spite of the line of reasoning developed above, it could be argued that the recognition of fractions involves a greater degree of calculation, and hence establishment of relationships, than the model presented would suggest. The model given assumed that the basic process was one of matching segment shapes to some kind of internalised templates, and that the main development in this type of recognition would involve the addition of new templates to the repertoire and tightening of the match-mismatch criterion. However, there are other possible models. For example, fraction recognition may involve the iteration of an internally generated size analogue symbol (s.a.s.) into the segment to be quantified. Thus the fraction three-quarters would be calculated by applying the s.a.s. for a quarter three times, whereas with the previous model three-quarters is recognised as a unit. Alternatively the smaller segment of the circle might be used as a metric for the larger, and the number of iterations necessary to cover the larger segment calculated.

The examples of misquantifications given tend to support the original model rather than the alternative 'iterative' models. However, this is not very conclusive and so it is important to keep in mind the possibility that fraction recognition may be more complicated than it appears. One possible test between the alternative models would be to measure the time taken by children to recognise different fractions. With the iterative models an increase in processing time proportional to the number of iterations necessary would be expected, whereas the model adopted here should not lend to differences of this kind.

Itis hoped that this descriptive treatment of the methods of quantification revealed by the pilot experiment will assist the reader in understanding the results of the next experiment. No figures have been quoted because a principle is being followed of using short, clinical, experiments in order to prepare the ground for and elucidate certain aspects of the more conventional experiments, and it is felt that the quotation of figures obtained in non-standardised situations is misleading, if not invalid.

The roulette experiment.

The number of different possible methods of quantifying data in this kind of situation revealed by the pilot experiment, and the different outcomes produced by the different methods, means that an experiment which is both logically identical to the first beads experiment for all subjects and standardised for all subjects cannot be designed. However, an experiment which will give a similar range of problems to the first beads experiment is feasible. The 'roulette' experiment is thus a compromise between this and its previously stated aims.

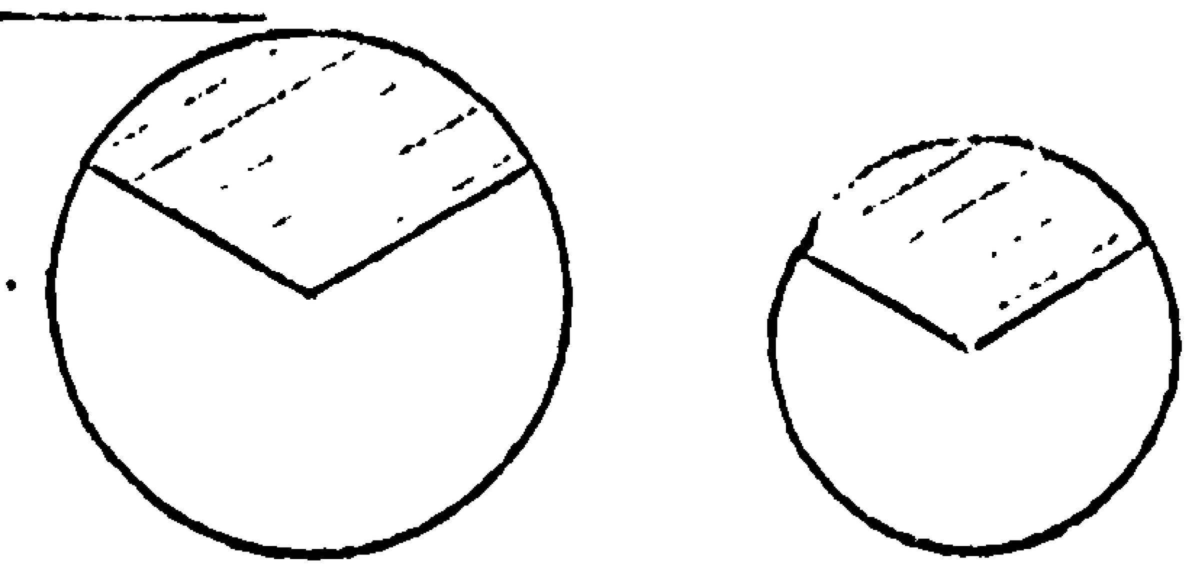
As in the beads experiments a forced-choice situation was used. Subjects were faced with a large and small wheel, each shaded with a red and green segment. A pointer was placed on each wheel and the subject was asked which pointer he would spin if he wanted it to finish pointing to red, and which if he wanted it to finish pointing to green. After each choice the gamble was made.

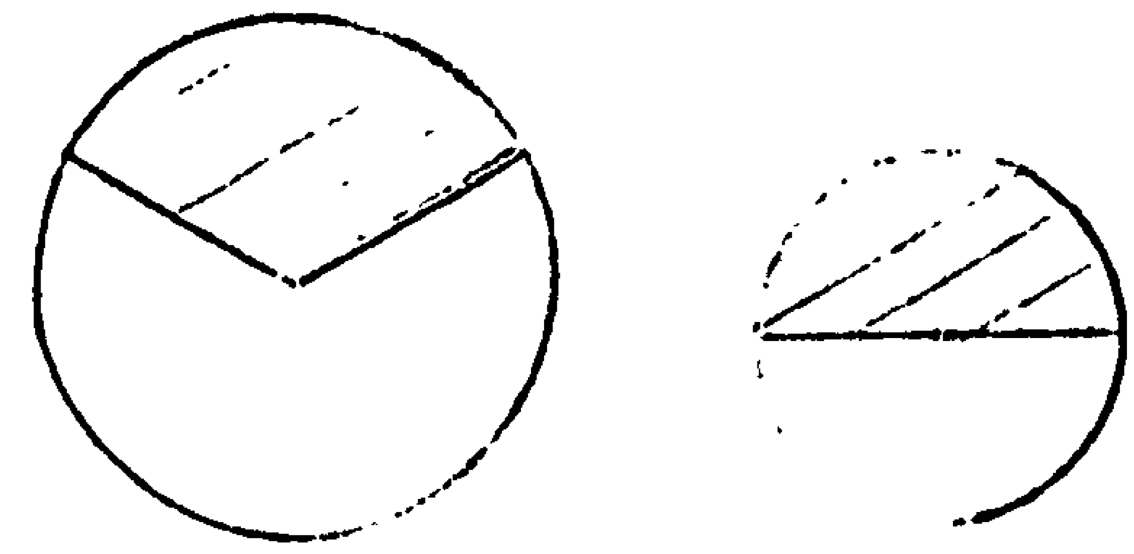
The fact that different methods of quantification can be applied to problems of this type means that the kind of design used in the first beads experiment, which is intended to identify certain strategies with a high degree of precision, is not possible here. Instead a design will be adopted for the main part of the roulette experiment which will exhaust all possible combinations of the fractions $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, of red and green on the large and small wheels. As long as colour preferences are ignored (so that a problem having a large wheel with $\frac{1}{3}$ red, $\frac{2}{3}$ green, and a small wheel with $\frac{1}{2}$ red, $\frac{1}{2}$ green, is regarded as equivalent to a problem having a large wheel with $\frac{2}{3}$ red, $\frac{1}{3}$ green, and a small wheel with $\frac{1}{2}$ red, $\frac{1}{2}$ green) this produces five problems. In addition, two extra problems were introduced where the discriminations required of the subject are finer ($\frac{2}{3}$ v $\frac{3}{4}$ and $\frac{1}{3}$ v $\frac{1}{4}$), these were given at the end of the experiment in cases where the experimenter considered that the result might be interesting. One of these problems was also given at the beginning of the experiment as a pretest to ensure that the subject had understood the instructions. Two pretest problems had been used in the beads experiments, but this was found to be excessively cautious in view of the ease with which the instructions seem to be understood.

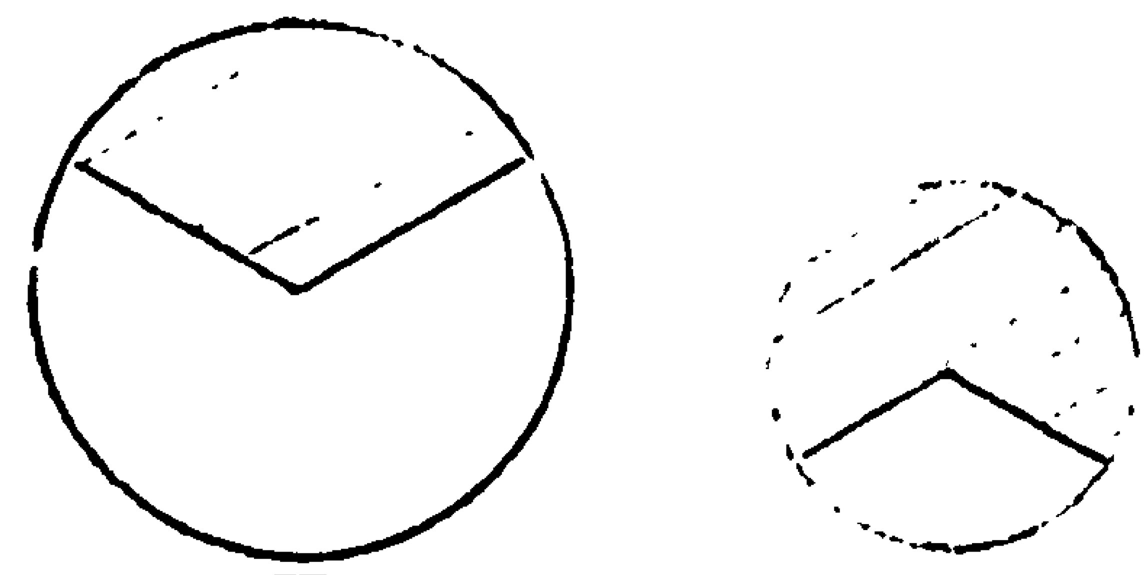
ABLE 5: PROBLEMS USED IN THE ROULETTE EXPERIMENT

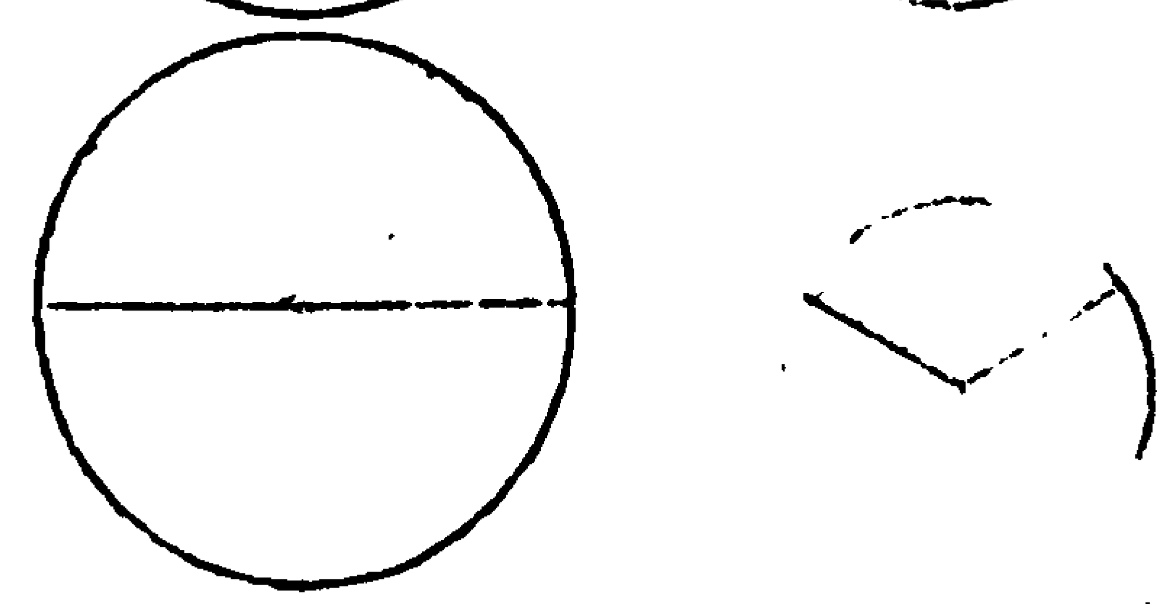
PRETEST: PROBLEM R6 OR R7.

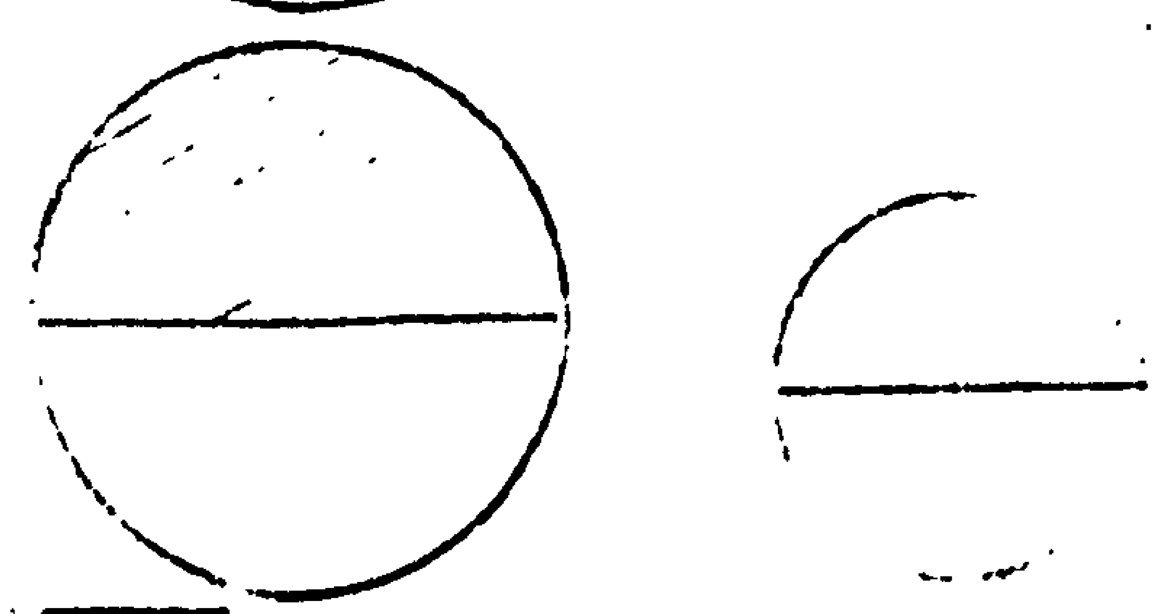
STUDY PROBLEMS:

R1: $\frac{1}{3}$RL v $\frac{1}{2}$RS i.e. 

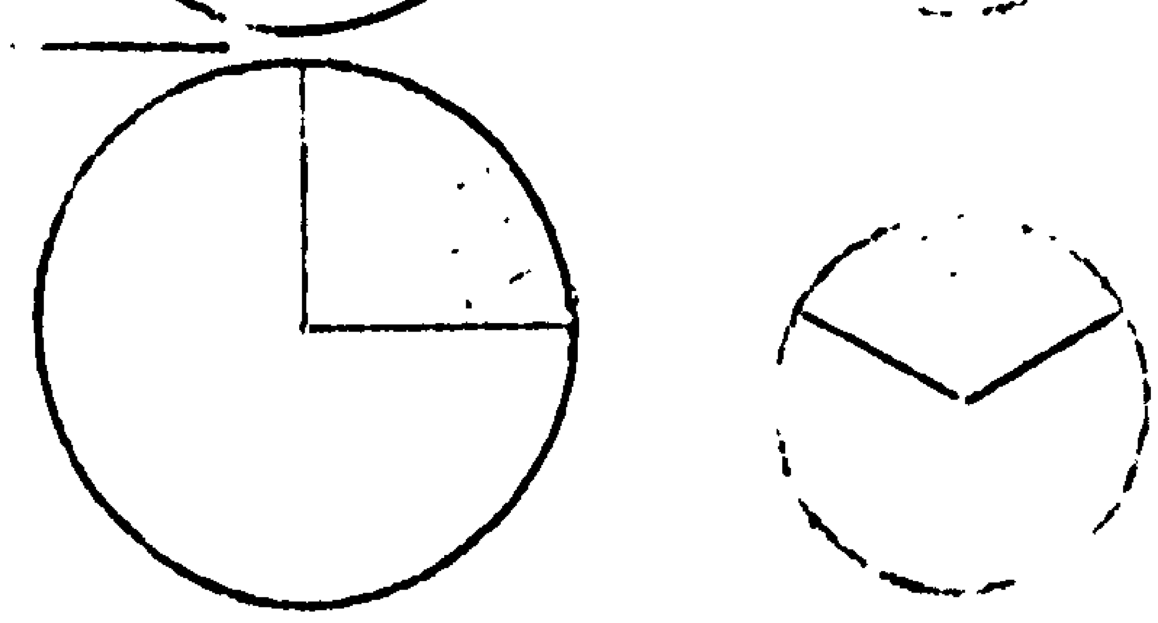
R2: $\frac{1}{3}$RL v $\frac{1}{2}$RS i.e. 

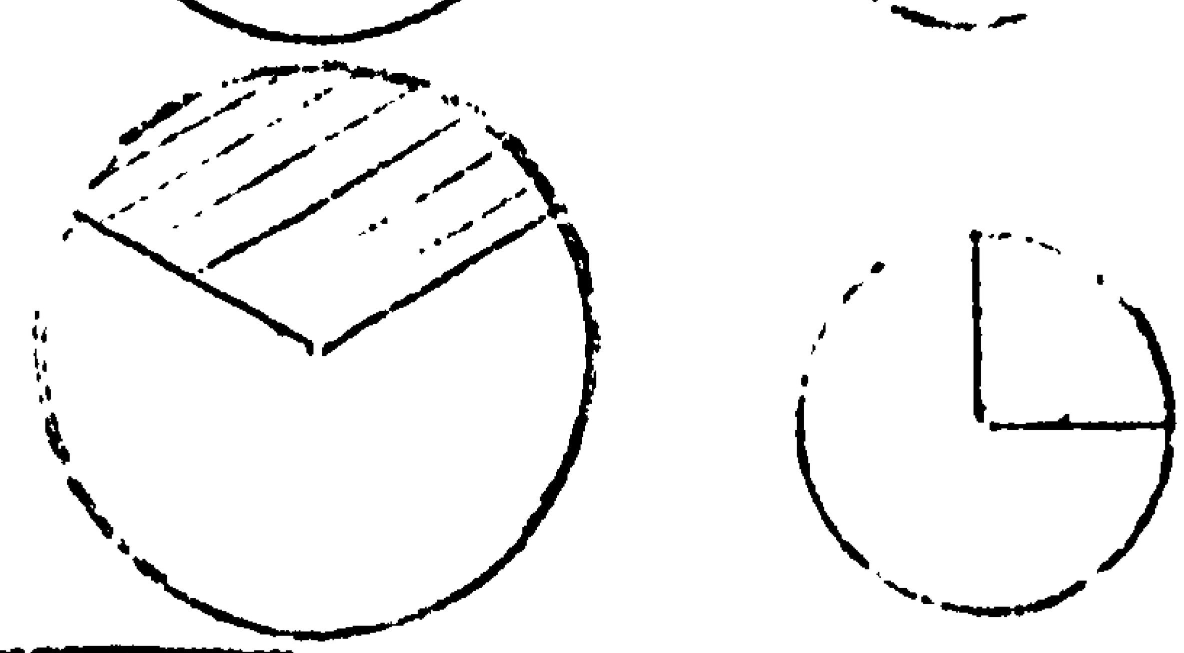
R3: $\frac{1}{3}$RL v $\frac{2}{3}$RS i.e. 

R4: $\frac{1}{2}$RL v $\frac{1}{3}$RS i.e. 

R5: $\frac{1}{2}$RL v $\frac{1}{2}$RS i.e. 

ADDITIONAL PROBLEMS:

R6: $\frac{1}{4}$RL v $\frac{1}{3}$RS i.e. 

R7: $\frac{1}{3}$RL v $\frac{1}{4}$RS i.e. 

THE EXPERIMENT CONSISTS OF THE PRETEST, FOLLOWED BY R1, R2, R3, R4, R5, IN RANDOMISED ORDER, THEN R6, R7, AT THE EXPERIMENTER'S DISCRETION.

KEY: $\frac{1}{2}$RL = LARGE WHEEL, HALF RED, HALF GREEN.

 = GREEN SEGMENT.  = RED SEGMENT.

Procedure: The general format and instructions of the experiment have been arranged to make comparison of the results obtained with the results of the beads experiments as easy as possible. The experiment begins with a pretest intended to make the subject familiar with the experimental situation, materials, and instructions. The pretest is followed by the five experimental problems, presented in an order varied from subject to subject by shuffling the experimental cards. Finally, there are the two additional problems, which replace the'clinical' interview often used in Genevan experiments, and which are only given to certain subjects. Each of the problems used in the experiment has two parts, namely the part where red is the target colour and the part where green is the target colour, which are always presented in this order (reasons for this, and its effect on the results, will be discussed in detail later). The orientation of the segments on the wheels and the left or right hand position of the large and small wheels were varied haphazardly throughout the experiment (again, by shuffling the experimental cards).

The materials used consist of two pointers which can be spun, and these are placed on cards with a wheel drawn on, The wheels are either 165 mm. in diameter or 280 mm. in diameter, with a red and green segment marked by paralled shading lines 13mm. apart. No representation of the situation of the kind used in the beads experiments was necessary as all the relevant features remain visible throughout the experiment. Data were collected by the experimenter in the form of a written record of the reasons given by the subject to justify his choices, the choices he makes, and their success or failure in producing the desired outcomes. The order of presentation of the problems was also recorded . The problems used are illustrated in Table 5.

The initial instruction given to each subject were as follows:

'We're going to try out some games and I want you to tell me what you think of them.

If we spin this pointer it can stop on the red bit of the circle or the green bit.' (This was then demonstrated with one of the pointers).

'And if we spin this one it can stop on the red bit or the green bit as well.' (This was demonstrated with the other pointer).

'What I want to know is if we wanted to get one to stop on a red bit would it be better to spin this one or this one?' (Pointing as appropriate).

'Now why did you choose that one rather than the other one?'

(If the reply to this is unclear, ask 'What do you mean by that?')

When an answer has been given, 'Now you can try spinning the one you chose and if it stops on the colour we want,you get a point, if it stops on the other colour you get nothing. If it stops on red you get a point, if it's a green you get nothing. O.K.?'

If the subject didn't seem to understand any part of these instructions, it was repeated. After the initial trial a shortened form of instruction could be used:

'Now this time we want to get one to stop on a green/red (as appropriate)bit,so which one would you spin now?

Why did you choose that one rather than the other one?'

(If the reply to this is unclear, 'What do you mean by that?')

When an answer has been given: 'Well, we'll try spinning the one you chose and if it stops on the colour we want you get a point, if it doesn't you get nothing. If it stops on a green/red you get a point, if it's a red/green you get nothing. O.K.?'

As in the beads experiments these instructions were memorised by the experimenter in order to allow the situation to be as natural as possible.

Sample: The experiment was carried out in the same primary school as the beads experiments. Seventy-two children were used, divided into six groups of twelve subjects aged 6 years, 7 years, 8 years, 9 years, 10 years, and eleven years. Some of the these subjects had already taken part in one or more of the beads experiments, but this was not considered important as the roulette experiment was carried out some nine months after the last of the beads experiments. Ten of the subjects come into this category, they are SR5 (SB6), SR16, (SB24), SR18 (SB21), SR22 (SB15), SR30 (SB25), SR33 (SB29), SR51 (SB48), SR58 (SB56), SR60 (SB57), and SR 69 (SB63). (The sizes of the classes in the primary school used for the study did not allow completely independent samples of children to be used in every experiment).

Results of the roulette experiment.

The results were analysed both as data in their own right and as data to be compared with the results of the first beads experiment. In order to provide a comparison with the first beads experiment results the categorisation scheme used for those results was applied to the results of the roulette experiment, and the validity of the scheme in this context was assessed independently by means of the Brimer cluster analysis.

In order to apply the categorisation scheme developed for the beads experiment it is necessary to specify criteria for deciding which of the four theoretical models a given response corresponds to. The following list was drawn up:

Type 1: Any answer corresponding to the first stage of development found in the results of the beads experiments. Such answers seem to the experimenter to involve spurious factors, or guesswork. Choices made are erratic and their justifications often seem post hoc. The following illustrations of this type of answer are taken from the results of the experiment:

SR11 (roulette experiment subject number 11), problem R5 green (problem number R5, target colour green.)
Chooses $\frac{1}{2}$RL (the large wheel, which has $\frac{1}{2}$ red and $\frac{1}{2}$ green marked on it). 'Every time it goes on the right one'.

SR11, problem R3 red.
Chooses $\frac{2}{3}$RS. 'I haven't tried it'.

SR12, problem R4 green.
Chooses $\frac{1}{2}$RL. 'It goes faster'.

SR14, problem R3 red.
Chooses $\frac{1}{3}$RL. 'It's got those bits there'. (Pointing to shading lines).

SR15, problem R2 green.
Chooses $\frac{1}{2}$RS. 'Don't really know why'.

SR15, problem R1 green.
Chooses $\frac{1}{3}$RL. 'I just tried the little one'.

SR19, problem R5 red.
Chooses $\frac{1}{2}$RS. 'It's a small circle'.

Type 2: Answers corresponding to the model 2 strategy. Such answers involve a comparison of the amount of each wheel covered by the target segment, and a choice of the wheel with the larger amount covered. Answers involving comparisons of the amount of each wheel covered by the non-target segment followed by choice of the one with less are also scored as type 2, although these are rare. The actual method used for the quantifications is disregarded as long as the subject indicates that he compared the amount of each wheel covered by the target segment and chose the one with more.

Examples of type 2 answers:

SR23, problem R4 red.

Chooses $\frac{1}{2}$RL. 'Got more red than that one'.

SR24, problem R5 red.

Chooses $\frac{1}{2}$RL. 'It's got a bigger space of red than that one. It's a lot better'.

SR32, problem R3 red.

Chooses $\frac{2}{3}$RS. 'It's three-quarters of red. Other has less'.

SR34, problem R3 red.

Chooses $\frac{2}{3}$RS. 'The red's three-quarters, only one quarter red on the other'.

SR36, problem R7 green.

Chooses $\frac{1}{4}$RL. 'The green goes round the circle more than on the other one'.

SR41, problem R5 red.

Chooses $\frac{1}{2}$RL. 'It's still bigger. This red half is bigger than that half a circle.'

SR44, problem R5 green.

Chooses $\frac{1}{2}$RS. 'It's just a guess. Looks the same amount on each one'.

SR45, problem R2 red.

Chooses $\frac{1}{3}$RS. 'There's too much green on the other one. More than on this one'.

SR47, problem R6 red.

Chooses $\frac{1}{4}$RL. 'It's got more red lines than the little one'.

SR50, problem R1 red.

Chooses $\frac{1}{3}$RL. 'That red bit's wider than the other one'.

Type 3: Answers indicating use of the model 3 strategy. This involves comparison of the difference between the amount of each wheel covered by the target and non-target segments, and choice of the most favourable case (i.e. the wheel with the largest 'surplus' or the smallest 'deficit'). As with the type 2 answer category the method of quantification used is disregarded.

Examples of type 3 answers:

SR24, problem R4 red.

Chooses $\frac{1}{2}$RL. 'They're the same red and green, the other has a smaller space of red than green.'

SR32, problem R4 green.

Chooses $\frac{1}{3}$RS. 'There's one quarter more green than red.'

SR47, problem R2 green.

Chooses $\frac{1}{3}$RS. 'There's more green than red lines'.

SR59, problem R7 green.

Chooses $\frac{1}{4}$GS. 'More green than red. Less red than the other one, a bit of green extra'.

SR60, problem R1 red.

Chooses $\frac{1}{3}$RL. 'Doesn't matter, green has more on both than red'.

SR72, problem R2 green.

Chooses $\frac{1}{3}$RL. 'There's more of it than the red, it goes over the half way marks'.

SR46, problem R3 red.

Chooses $\frac{1}{3}$RS. 'Got more red on than green'.

It must be pointed out that some of these answers only refer to one wheel, so that they do not completely meet the criteria for the type 3 categorisation. However, as all the responses of this type produced in the experiment were found to be associated with choices consistent with the model 3 strategy, they were classed as type 3, since the impression is gained that the lack of reference to the other wheel is due to 'laziness' rather than not having taken it into account. It is interesting to note that the same high consistency was not found in a later experiment.

<u>Type 4</u>: Answers which seem to involve some understanding of proportional relationships. For example:

SR51, problem R1 red.

Chooses $\frac{1}{3}$RL. 'They're the same. If you make that one smaller it's the same'.

SR72, problem R1 red.

Chooses $\frac{1}{3}$RL. 'Any one. They're both the same if you cut them down to their sizes. I'll have the big one'.

<u>Type '?'</u>: As in the beads experiments, a certain number of responses were found to be unclassifiable within this scheme, and were grouped together as type '?'. All of these responses appear to involve application of the model 2 or model 3 strategy with a choice of the wheel with less rather than more of whatever is required. Examples are:

SR22, problem R5 green.

Chooses $\frac{1}{2}$RS. 'Got less green than that one'.

SR22, problem R4 green.

Chooses $\frac{1}{2}$RL. 'Got less green than that one. This is straight, that goes up'. (A reference to the shapes of the boundaries between the red and green segments).

SR38, problem R2 red.

Chooses $\frac{1}{3}$RL. 'The red's smaller than the green'.

It is perhaps as well to point out the ways in which this classification scheme has had to be modified from the scheme used in interpreting the results of the beads experiments. The main difference is caused by the fact that in the beads experiments the almost universal use of numbers as quantitative symbols means that the choices to each part of each problem which would be consistent with each strategy can be specified precisely, so that if a subject does not choose in accordance with the reason he gives we can ponder about his 'real' reason and, more importantly, the fact that most choices and reasons are not inconsistent supports this type of analysis.

Such clear-cut decisions cannot always be made in the analysis of the roulette results because of the highly personal nature of estimation and its size analogue symbols. If a subject maintains that half a small circle seems bigger thana a third of a larger circle the experimenter is in no position to contradict him, and likewise if he had maintained that half the small circle seemed smaller than a third of the larger one. The validity of an analysis of the reasons for choices obtained in the roulette experiment might thus be questioned were it not for the support given to this method by the results of the beads experiments.

Another difference is that in the classification scheme for the roulette results the criteria for identifying an answer as belonging to type 4 have been relaxed. In the beads experiments type 4 was used to refer to answers which seemed to be generated by the model 4 strategy. It would not be possible to do this with the results of the roulette experiment because accurate recognition of fractions removes the need for most of the calculation involved in the model 4 strategy. Because of this it was decided that it would be most profitable to use the type 4 category as a way of distinguishing the subjects who appeared to show some genuine grasp of proportion from those subjects who merely attached a fractional label to a certain range of segment shapes. This obviously involves a greater than usual degree of interpretation by the experimenter and it is left to the reader to decide whether he thinks that this is justified by the examples given. The category turns out to be comparatively rare in the results anyway, so that the decision as to its validity is not as important as it might have been.

Even though the type 4 category is uncommon in the roulette results it could be held that the relaxation of its defining attributes damages the comparison about to be made, between the results of the beads and roulette experiments. In order to find out whether or not this is the case the results of the first beads experiment were re-interpreted according to the new set of type 4 criteria, and no difference to the original interpretation was found. In other words, the beads experiment provides the same examples of the type 4 category whether the strict or relaxed criterion is used, whereas the roulette experiment has some examples of type 4 according to the relaxed criterion, but none by the original strict criterion.

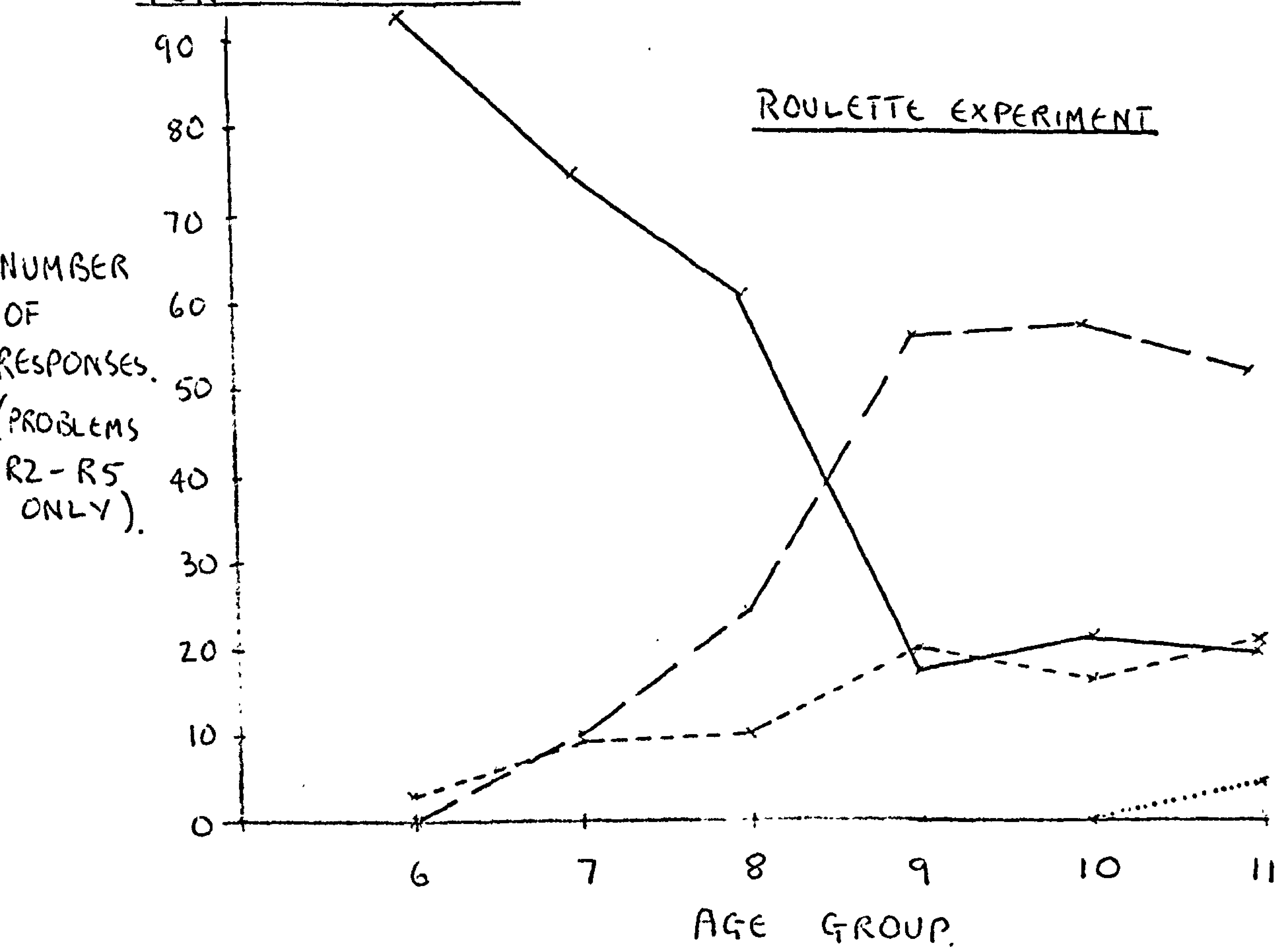
The result of applying this classification scheme to the data collected in the roulette experiment is set out in Appendix E. A graphical illustration can be produced in order to facilitate comparison with the result of the first beads experiment. There are eight responses from each subject in the first beads experiment, and ten per subject in the roulette experiment, whilst the number of subjects of each age sampled is the same in each experiment, so that the comparison could be made by multiplying by 4/5 the scores in the table of frequency of response types against age for the roulette experiment (see Appendix E, the comparisons could also be effected by means of proportions, but multiplying by 4/5ths has the same effect, and is more convenient). However, the comparison will only be meaningful if the sets of problems used in the two experiments can be seen as in some way equivalent. It has been pointed out that it is impossible to make the underlying structure of the two sets of problems identical for all subjects, but that a similar range of problems in both cases can be achieved. Because of this, the design of the roulette experiment was worked out with the aim of providing an experiment which would be satisfying in its own right as well as a possible comparison with the beads experiments. This leads to the roulette experiment's having two of the main experimental problems of the 'tricky' kind where each collection has the same proportion of red and green, and the subject must point out their equivalence rather than choosing one (to pass). In the first beads experiment only one of the main experimental problems was of this form.

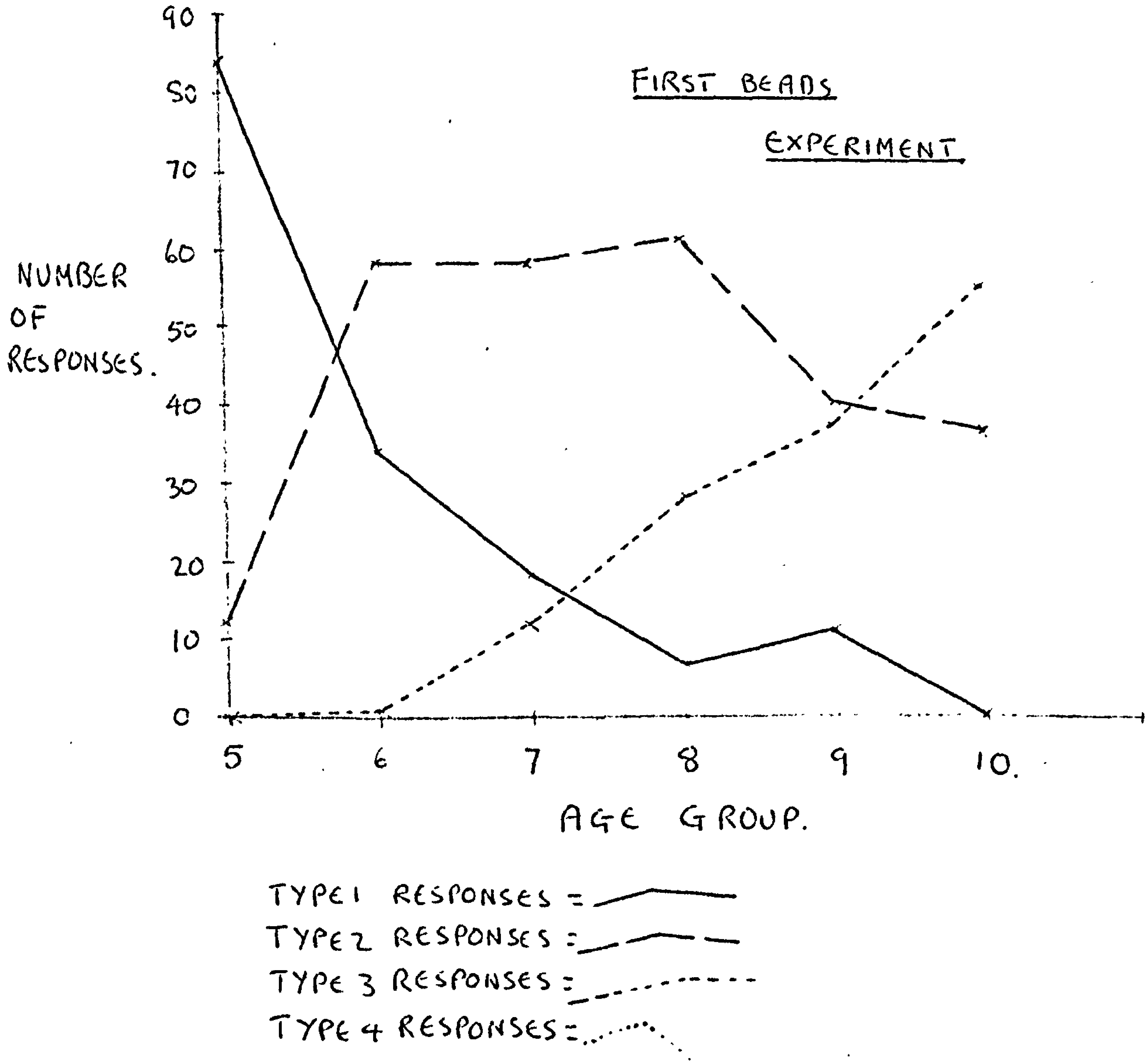
The problems referred to are:

	Collection A	Collection B
Beads problem B3:	4 red and 4 green beads	2 red and 2 green beads
Roulette problem R1:	Large wheel $\frac{1}{3}$ red and $\frac{2}{3}$ green.	Small wheel $\frac{1}{3}$ red and $\frac{2}{3}$ green.
Roulette problem R5:	Large wheel $\frac{1}{2}$ red and $\frac{1}{2}$ green.	Small wheel $\frac{1}{2}$ red and $\frac{1}{2}$ green.

The presence of two problems of this kind in the roulette experiment

FIGURE 10: GRAPHICAL ILLUSTRATION OF THE RESULT OF THE ROULETTE EXPERIMENT IN TERMS OF THE SOLUTION STRATEGIES USED AT DIFFERENT AGES, TOGETHER WITH AN EQUIVALENT GRAPH OF THE RESULT OF THE FIRST BEADS EXPERIMENT FOR COMPARISON.





and only one in the first beads experiment constitutes a source of difference between the two sets of problems, so that a more meaningful comparison might be obtained if one of the roulette problems is ignored. The best one to ignore seems to be problem R1, since problem R5 and beads problem B3 both utilise the same proportions of red and green for their collections. The trouble with doing this is that whether the results of problem R1 are ignored or not the problem was present in the experiment and may influence the results obtained from the other problems in some significant way. This seems unlikely, but the question is not of pressing importance since Appendix E includes tables of the frequencies of the different types of response at different ages compiled by including all problems and multiplying each result by four/fifths, and by only using results from problems R2 to R5, which demonstrate that there is little difference between the two methods except in the number of type 4 responses, which goes down if only problems R2 to R5 are considered. As the type 4 responses are only produced on problems R1 and R5, the results from problems R2 to R5 only have been used in the comparison of the roulette and first beads experiment results which is illustrated by figure 10.

In order to test the difference between the proportions of stage 1, 2, and 3 answers obtained in the first beads and roulette experiments, the chi square test for two independent samples was used,with subjects in the age range six to ten years, (strictly speaking, the samples are not completely indpependent, but in view of the very different format of the two experiments and the long time interval between them the use of the test was considered justifiable). In order to apply the test each subject was assigned to the stage to which most of his answers belonged (ties were resolved by assigning the subject to the lower of the tied stages). The difference between the proportions of subjects falling into the three stages in each experiment is significant at the .001 level (See appendix J).

Discussion: From the graphs of the results of the first beads and roulette experiments (figure 10) it can be seen that there are two main differences between them. Firstly, the results of the roulette experiment appear retarded by comparison with the results of the beads experiment, if we consider that the four response types represent increasingly relevant and sophisticated strategies. Secondly, the model 3 strategy (type 3 answers) features less prominently in the results of the roulette experiment, although it is still quite clearly present. A minor difference is the appearance of a few type 4 answers at a much younger age in the roulette than the beads experiment, but this is of questionable significance. The shapes of the curves representing the type 1 answers and the curves representing the type 3 answers obtained in each experiment from subjects between six and ten years of age are closely comparable, and median tests show that the median of the type 1 curves is significantly higher in the roulette experiment, whilst the median of the type 3 curves is significantly higher in the first beads experiment. (See appendix J.)

The apparent drop in performance between the first beads and roulette experiments is consistent with the prediction made that the unreliability of Q_e would cause the results of the roulette experiment to appear retarded. However, this is by no means the full story. The results of the roulette experiment show much more heterogeneity than the results of the first beads experiment and it is only by virtue of the conceptual framework which has been developed that much sense can be made of them.

In the beads experiments any quantitative symbols generated to aid solution of the problems were almost invariably numerical, so that the way in which the quantitative symbols were organised to arrive at a solution provided the major source of variability between answers. In the roulette experiment, on the other hand, although Q_e is the principal quantification operator used, there are a variety of ways in which it can be applied, so that the way in which the quantitative symbols it produces are organised can only account for some of the variance in the results and is no longer so illuminating. If the examples of answers corresponding to each of the major strategies in the beads and roulette experiments are compared a striking difference can be seen between the homogeneity of the examples of these categories in the beads experiment and the heterogeneity of the roulette experiment examples.

This point about the role of the different types of quantitative symbol as well as different strategies involved in the roulette performances merits further discussion. Suppose that the model 2 strategy is used to solve problems of the type involved in the roulette experiment. In order to apply this strategy it is necessary to generate quantitative symbols representing the amount of the target colour which is marked on each wheel. If this is done by estimating the area occupied by the target segment on each wheel the choice finally arrived at will sometimes be the wheel with the highest mathematical probability of success, and sometimes the other wheel. If, on the other hand, the shapes of the target segments are recognised as corresponding to certain fractions, and these fractions are then used as quantitative symbols, the choice made will always be the wheel with the highest mathematical probability of success, as long as the fractions have either been recognised correctly or been adequately discriminated in the range necessary for correct solution of the problem. This is caused by the fact that the idea of proportion, so necessary to the mathematical conception of probability, is implicit in fractions, which are expressed as parts of the whole object. In fact, fractions express the ratio of part to whole which is required for the calculation of the mathematical probability, rather than the ratio of part to part, which would have to be turned into an expression of part to whole so that the probability of success can be calculated.

Despite this, errors (in the sense of deviation from the choice which has the highest mathematical probability of success) are possible when fractions are used as quantitative symbols together with the model 2 strategy, because there is no method available in the situation used for calculating these fractions precisely, they must simply be recognised. The level of discrimination needed to get a 'correct' solution by this method varies from problem to problem. In problems R6 and R7 it is necessary to spot the difference between $\frac{1}{3}$ and $\frac{1}{4}$, or $\frac{2}{3}$ and $\frac{1}{2}$, and to know in each case which is larger. In the main experimental problems, R1-R5, it is only necessary to know the relationships between $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$, but the point is that the source of error now lies in the quantification method itself rather than the way the quantified information is used.

This property of fractional quantitative symbols, that they encapsulate the relation of part to whole, may be one reason why the model 3 strategy is not found so often in the roulette experiment as in the beads experiments, since anyone with some understanding of fractions will see that it has been superseded. Another possible reason concerns the fact that the model 3 strategy requires the comparison of two pairs of quantitative symbols, followed by a comparison of the results of these comparisons. The size analogue symbols arising from area estimation will often lack the precision necessary for this to be possible, except in cases where the comparisons are of a gross order. The point at which a particular child will simply produce nonsense from attempting to use the model 3 strategy will obviously depend on the point to which he has managed to refine his use of estimation of areas. A similar point can be made concerning the use of the model 3 strategy with quantitative symbols arrived at by recognising fractions.

A further complication caused by the depth of relationship implicit in fractions is that this leads to subjects' verbal reports being less informative than in the beads experiments. When a subject gives a justification for a choice involving the model 2 strategy with fractions as quantitative symbols he may understand quite a lot about the relevance of the relation of part to whole in this kind of situation, or he may understand very little. He may take for granted the experiment's understanding of the implications of the fractions, or he may not understand them himself. Without introducing extra and probably non-standardised questioning into the experiment it is almost impossible to tell which is the case. Studies of the way in which children solve problems involving proportionality have been reported by Lunzer and Pumfrey (1966) and Pumfrey (1968), but it would be unwise to make inferences from these to the present situation.

As well as the variety of methods of quantification which can be seen in the results of the roulette experiment, the results of the categorisation of each subject's responses into four types also reveals a more heterogeneous result than the beads experiments. In the main part of the beads experiments several subjects gave only responses corresponding to stage 1, stage 2, or stage 3, and the other subjects only displayed mixtures of adjacent stages (ignoring unclassifiable answers). In the roulette experiment 26 subjects gave all type 1 responses, but only 1 subject gave all type 2, and 1 subject all type 3. Furthermore, if a developmental trend from responses of type 1 to responses of type 2 to type 3 to type 4 is assumed, many of the subjects' profiles show mixtures of more than just adjacent response stages. There are cases of mixture of type 1 and type 3, of type 1, type 2, and type 3, of type 2, 3, and 4, of type 1, 2, and 4, and in one case of type 1, 2, 3, and 4, as well as the adjacent mixtures of types 1 and 2, and types 2 and 3. In other words, the only 'pure' stage which seems to exist corresponds to the type 1 responses, and since type 1 is more of a ragbag of irrelevant strategies than a unified category this isn't a very promising start.

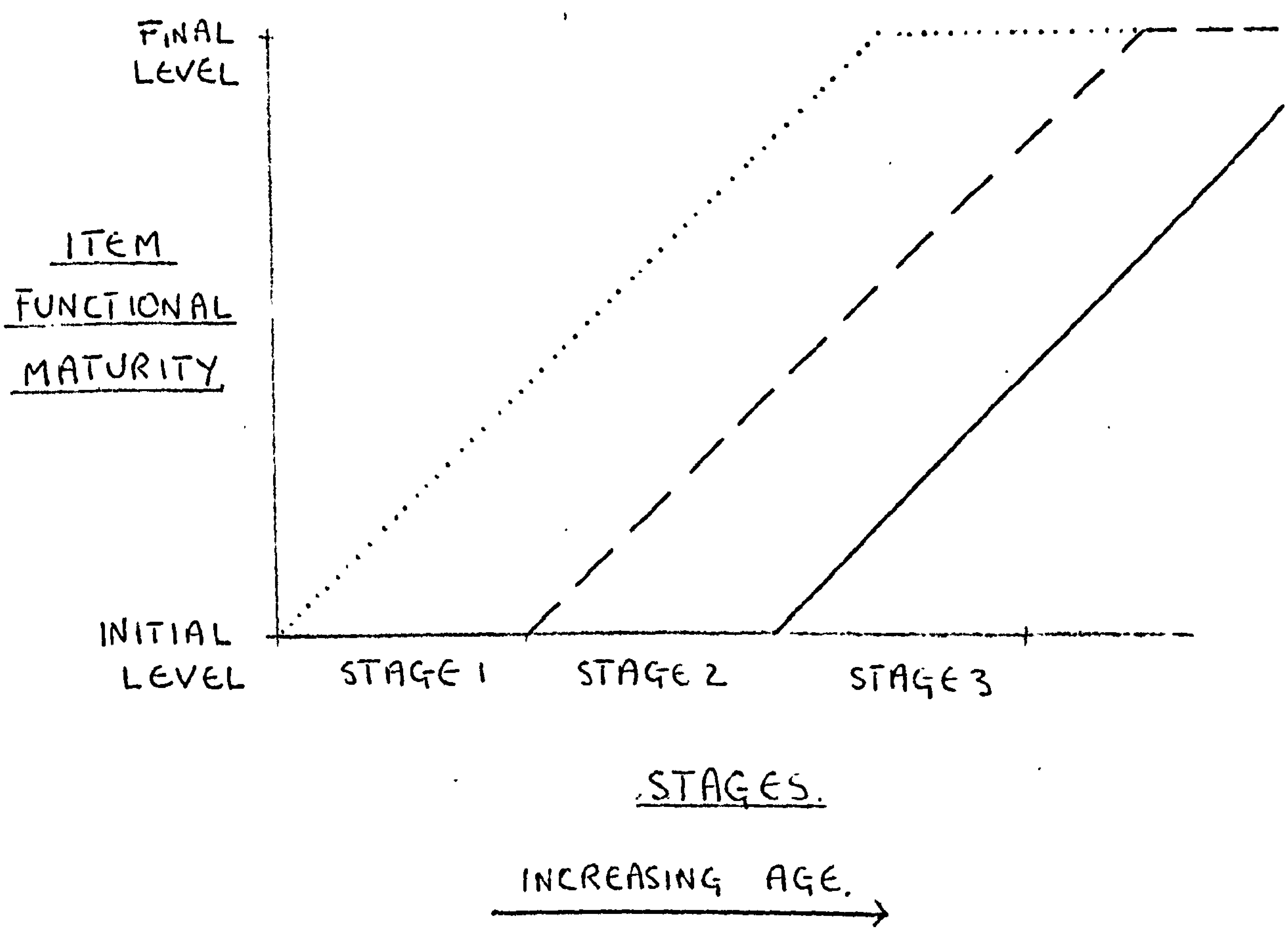
There is some evidence supporting a stage interpretation of the results of the roulette experiment in the graph of these results. This shows that around 8 or 9 years of age the type 2 responses take over as the dominant type of response from the type 1 responses. The use of model 3 (type 3 responses) increases throughout the age range sampled (up to 11 years old), but never takes over from model 2 (type 2 responses) in the way that it does in the beads experiment. The number of type 4 responses also increases throughout the age range, but can hardly be said to be important.

One noticeable feature of the result of the roulette experiment is

FIGUR

FIGURE 11. A MODEL OF A STAGE DEVELOPMENT CONSISTENT WITH THE RESULT OF THE ROULETTE EXPERIMENT.

NUMBE
OF
RESPON
(PROBL
R2 - R
ONL



NUM
OF
RESP

that the slopes of the curves contained in the graph of those results are less steep than the slopes in the corresponding graph of the results of the first beads experiment. This is perhaps significant as the roulette results form the sort of pattern that might be expected from a version of what Flavell (1971) calls the 'gradual development model'. The results of the first beads experiment correspond to a version of Flavell's (1971) developmental model 1b (see figs. 5,8,9), in which each stage is exhibited in a pure form for a time followed by a period of transition involving stage mixture of the stages preceding and following this transition. By using the criterion of stage mixture adapted from Turiel (1969), it is apparent that the results of the roulette experiment show the more gradual stage development of Flavell's (1971) model 1c (see figs. 5, 8). A version of this model which has been modified in the same manner as the model adapted from Flavell (1971) for the interpretation of the results of the beads experiments is shown in figure 11.

Cluster analysis of the results of the roulette experiment.

In order to obtain an independent check on the validity of the interpretation of the results of the roulette experiment put forward the Brimer cluster analysis was again employed. The analysis was applied to two sets of results:

Analysis A; the results from problems R2 - R5 only. This analysis was performed to facilitate comparison of the results of the roulette and beads experiments.

Analysis B; the results from all the experimental problems, R1 - R5. This analysis was performed in order to make an appraisal of the overall result of the roulette experiment in its own right.

The underlying theory and general method employed in the Brimer analysis have already been described, so the comments made here will be restricted to the details necessary to an understanding of the way in which it was carried out in this instance.

The first step in the analysis is the categorisation of all the responses obtained to the different items used in the experiment into categories consisting of indistinguishable responses to the same item. The most convenient way of describing this categorisation is to begin by outlining all the categories found for one item, and then adding any further categories required for other items to this list. As problems R2, R3, and R4 require a rather different categorisation scheme to problems R5 and R1, and as R1 is only used in one of the two analyses, the first item to be considered will be the part of problem R2 where red is the target colour.

Problem R2 involves a large wheel which is divided into a red segment covering one third and a green segment covering two thirds of its surface, and a small wheel which is half red and half green (this situation will be represented by the convention $\frac{1}{3}$RL v $\frac{1}{2}$RS). When red is the target colour a pass choice is defined as choice of the smaller wheel ($\frac{1}{2}$RS) and a fail choice is choice of the larger wheel ($\frac{1}{3}$RL). The following response categories could be distinguished on this item:

FU: A 'fail' choice without any reason given, or an unqualified assertion as to the correctness of the choice, or repetition of the experiment's instruction.

e.g. SR6: chooses $\frac{1}{3}$RL. 'Don't know why'.

PU: A 'pass' choice accompanied by a reason like the reasons specified for FU.

e.g. SR23: chooses $\frac{1}{2}$RS. 'I don't know why'.

SR43: chooses $\frac{1}{2}$RS. 'You can get to it better'.

FSR: A 'fail' choice accompanied by a 'spurious' reason, such as the size of the wheel chosen, or the speed at which the pointer can rotate.

e.g. SR4: chooses $\frac{1}{3}$RL. 'It's biggest'.

PSR: A 'pass' choice with a reason like FSR.

e.g. SR8: chooses $\frac{1}{2}$RS. 'It's the littlest'.

SR1: chooses $\frac{1}{2}$RS. 'It'll go much faster'.

These examples seem at first sight to belong to two categories, namely choices of the small wheel, and assertions that one wheel is faster than the other. However, inspection of all the results of this type shows that they cannot be reliably separated in this way, because many of the answers are like this:

SR19: chooses $\frac{1}{2}$RS. 'It's small. It'll go slow'.

i.e. many children seem to equate the small size wheel with slow speed of rotation of the pointer.

As the answers cannot be reliably separated into further categories they are left together.

FPR: A 'fail' choice with a reason based on previous outcomes or previous choices, e.g. left-right alternation, large-small alternation, win-stay lose-shift strategy.

e.g. SR11: chooses $\frac{1}{3}$RL. 'I haven't tried it.'

PPR: A 'pass' choice with a reason like FPR.

e.g. SR7: chooses $\frac{1}{2}$RS. 'I got green before'.

SR15: chooses $\frac{1}{2}$RS. 'I haven't tried it'.

FRF: A 'fail' choice accompanied by reference to some relevant factor, without any clear reason for introducing it.

e.g. SR22: chooses $\frac{1}{3}$RL. 'The other's a half, this is not like a half'.

SR28: chooses $\frac{1}{3}$RL. 'It's a triangle shape, that is a rectangle.'

PRF: A 'pass' choice with a reason like FRF.

e.g. SR14: chooses $\frac{1}{2}$RS. 'It's got a half'.

PQS: A 'pass' choice accompanied by quantification of spurious factors.

e.g. SR12: chooses $\frac{1}{2}$RS. 'It's the same amount of lines on the bits'.

FQTE: A 'fail' choice with a reason based on a comparison of the sizes of the target segments on each wheel.

e.g. SR36: chooses $\frac{1}{3}$RL. 'The red's got a larger space than the other one'.

FQTW: A 'fail' choice with a reason based on comparison of the widths of the target segments on each wheel.

e.g. SR50: chooses $\frac{1}{3}$RL. 'It's wider than the other red bit'.

PQTE: A 'pass' choice with a reason based on a comparison of the sizes of the target segments on each wheel.

e.g. SR40: chooses $\frac{1}{2}$RS. 'It's got a bigger bit than that one of red'.

PQTF: A 'pass' choice with a reason based on a comparison of the fraction of each wheel covered by the target segment. The fractions will not necessarily be recognised accurately.

e.g. SR69: chooses $\frac{1}{2}$RS. 'Red's half of this circle, the other red one isn't'.

PQNTE: A 'pass' choice with a reason based on a comparison of the sizes of the non-target segments on each wheel, and choice of the wheel with the smaller non-target segment.

e.g. SR45: chooses $\frac{1}{2}$RS. 'There's too much green on the other one. More than on this one'.

PQNTF: A 'pass' choice with a reason based on a comparison of the fraction of each wheel covered by the non-target segment, and choice of the wheel with the smaller fraction of non-target segment.

e.g. SR71: chooses $\frac{1}{2}$RS. 'Half here. The other has more against it'.

PQRE: A 'pass' choice with a reason based on a comparison of the sizes of the target and non-target segments on each wheel.

e.g. SR24: chooses $\frac{1}{2}$RS. 'Same amount of red and green. More chance. The other has a smaller space of red than green.'

PQRF: A 'pass' choice with a reason based on a comparison of the fraction of each wheel covered by the target and non-target segments. The fractions are not necessarily recognised accurately.

e.g. SR60: chooses $\frac{1}{2}$RS. 'Halves. Got an even chance with green. Green has more space than red on the other'.

PQROE: A 'pass' choice with a reason based on a comparison of the relative sizes of the target and non-target segments on the wheel not chosen.

e.g. SR44: chooses $\frac{1}{2}$RS. 'That other one might get green and not red. More green on it than red.'

This completes the list of categories necessary to classify the responses obtained to the section of problem R2 involving red as the target colour. Many of the results to the section of R2 involving green as the target colour fall into the same categories, but the following additional categories are necessary:

FQTF: A 'fail' choice (choice of $\frac{1}{2}$RS) with a reason based on a comparison of the fraction of each wheel covered by the target colour. The fractions are not necessarily recognised accurately.

e.g. SR51: chooses $\frac{1}{2}$RS. 'It's got half green. The other green isn't half'.

PQREL: A 'pass' choice (choice of $\frac{2}{3}$RL) accompanied by a reason based on comparison of the numbers of lines on the target and non-target segments, without overt counting.

e.g. SR47: chooses $\frac{2}{3}$RL. 'There's more green than red lines. That one is the same red and green lines'.

PQRO: A 'pass' choice accompanied by comparison of the amount of target and non-target colour on the wheel chosen only. No reference is made to the other wheel.

This category is the only one in the analysis which is not defined in isolation since it was decided that if a subject produces a number of PQRO and PQRE answers, and none which might be FQRO, his PQRO answers will be considered as PQRE. This was done because the answer seem like either a lazy way of giving a PQRE justification, or a post hoc reason to support a guess, and it was felt important to keep these two possibilities apart. This is strictly not allowed by the analysis, but the choice is between tampering with this one category or deliberately allowing a possible source of artifact into the statistical analysis.

e.g. SR12: chooses $\frac{1}{2}$RL. 'That's smaller (points to the red segment) and that's bigger' (points to the green segment).

This completes the list of categories necessary for problem R2. The same categorisation scheme was found to cover problems R3 and R4 adequately, but modifications are necessary for problem R5. In problem R5 the large wheel is half covered by the red segment and half covered by the green segment, the small wheel is likewise half red and half green. This means that in both parts of the problem a pass response is defined as an assertion that it doesn't matter which wheel is chosen, and any preference for one wheel or the other would be regarded as a fail. The simplest way to accomodate the categorisation scheme used so far to the results of problem R5 is to add the suffixes B and S, representing choice of the large or small wheel, to the fail categories. Category FU would then become FUB or FUS, depending on which choice was accompanied by the FU reason. After the categorisation scheme has been modified in this way only two other categories are necessary to cope with R5. These will be described for the case where red is the target colour:

FQSB: A'fail' choice of the larger wheel with a reason involving quantification of spurious factors.

e.g. SR28: chooses $\frac{1}{2}$RL. 'It's got eleven lines, that has seven on the circle'. (Trying to count the number of shading lines on each circle).

PQRV: A 'pass' choice accompanied by a reason to the effect that the circles 'look the same'.

e.g. SR70: 'Both the same. They look the same'. Chooses $\frac{1}{2}$RS.
SR66: chooses $\frac{1}{2}$RS. 'I might be lucky. They're both the same except for the size. Lines in the same place.'

These two examples are not identical, because the second one is more explicit about what is meant than the first one. In the initial analysis the first example would be scored as type 2, whereas the second would be type 4. However, such a distinction rests heavily on the experimenter's interpretation of what the subject meant by what he said, and cannot be allowed into the Brimer analysis, where any distinctions made must be based on clear and firm criteria. As there are various gradations of explicitness between the answers brought together in this category no further reliable separation can be made and the category has not been broken up into smaller categories.

This completes the specification of the categories involved in the first of the cluster analyses of the roulette experiment results, analysis A, which involves problems R2 - R5 only. The second cluster analysis, analysis B, uses the results of all the main problems in the roulette experiment, R1 - R5. In problem R1 the red segment covers one third and the green segment two-thirds of each wheel. A pass response is thus an assertion that either wheel may be chosen, and preference for one of the wheels is regarded as a fail. This situation can be dealt with to a large extent by using the same response categories as those specified for problem R5, but the following additional categories are necessary for the part of problem R1 involving green as the target colour:

FQROB: A 'fail' choice of the **large** wheel with a reason indicating comparison of the amount of target and non-target colour on the wheel chosen only. No reference is made to the other wheel. e.g. SR44: chooses $\frac{1}{3}$RL. 'The green's bigger than the red'.

FQROS: A 'fail' choice of the small wheel with a reason like that outlined for FQROB.

e.g. SR54: chooses $\frac{1}{3}$RS. 'It's got more green than the red bit'.

FQRES: A 'fail' choice of the **small** wheel with a reason indicating an attempt to compare the relative amount of target and non-target colour on each wheel.

e.g. SR69: chooses $\frac{1}{3}$RS. 'It's bigger than the red. The red in the big **circle**'s too big for the green bit'.

With the addition of these categories the results to the part of problem R1 which has red as the target colour can also be incorporated into the scheme, and the description of the categories necessary for the Brimer analysis is complete.

Having completed the description of the categorisation scheme, there are certain aspects of it which must be clarified. The scheme is intended mainly as an aid to the reader, to show him the wide variety of answers produced in the experiment and to provide a way of making sense of the bewildering number of categories required. Because of this it is entirely post hoc. The way the categories were produced was simply by grouping together indistinguishable responses to the same experimental item, and the labels were applied to the categories after this had been done. This is more satisfying than fitting the results to a predetermined scheme, no matter how elaborate that scheme might be. Defining the categories only after they have been formed instead of starting with the category definitions leads to some important differences between the cluster and interpretative analyses, which have already been pointed out. What happens is that certain distinctions which can be made on a prior grounds turn out to be unworkable in the Brimer analysis because answers which may appear quite different are so intermixed in the results that they cannot be reliably separated.

The complete categorisation of the results of the roulette experiment can be found in Appendix F. When every category produced in response to each item of the experiment is assigned a separate number there are 116 categories for the analysis of R2-R5, and 148 categories for the analysis of R1 - R5. There are 72 subjects, so the data for the first analysis are within the capacity of the program used, but in the second analysis the number of categories must be reduced to 140 by the inclusion of 8 categories in other categories. This was effected by eliminating small categories which seemed very similar to other categories in the following manner:

Problem R1, red target: FQTW (2 instances) is included in FQTE.

Problem R2, red target: FQTW (1 instance) is included in FQTE.

PQS (1 instance) is included in PSR.

PQROE (2 instances) is included in PQRE.

Problem R2, green target: PQREL (1 instance) is included in PQRE.

PQRO (1 instance) is included in PQRE.

Problem R3, red target: PQRO (1 instance) is included in PQRE.

Problem R3, green target: PQRO (1 instance) is included in PQRE.

Thus, in the second analysis, the separation of PQRO and PQRE, which is of questionable validity as has already been pointed out, is abandoned. The results of the cluster analyses will now be presented in the same way as the results of the bead analyses were presented, starting with analysis A. The category numbering scheme and the list of category numbers for each subject, which constitutes the data for the cluster analysis program, is included in Appendix F.

Analysis A used data from problems R2 - R5 only, and generated the following first-order groups of subjects:

Group 1: A group of subjects who seem to solve the problems by choosing the wheel with the larger sized segment of the target colour. The PQTE and FQTE categories correspond to this method. The group has 23 members.

Group 2: A group of subjects who choose erratically and give reasons for their choices which are spurious or irrelevant by adult standards. The group has 15 members.

Group 3: A group of subjects who seem to solve the problems successfully by comparing the relative sizes of the target and non-target segments on each wheel, especially in response to Problem R3. The group has 16 members.

Group 4: A similar group to group 2.
Group 2 (15 members) and group 4 (13 members) have 12 members in common.

Group 5: A similar group to group 2 and group 4.
Group 2 (15 members) and group 5 (12 members) have 10 members in common.
Group 4 (13 members) and group 5 have 12 members in common.

Group 6: A similar group to group 1, but group 1 is made up of PQTE and FQTE categories, whereas group 6 is biased toward FQTE, especially in response to problem R5.
Group 1 (23 members) and group 6 (16 members) have 14 members in common.

Group 7: This group covers a wider range of responses than most of the other groups, and it is difficult to see a unifying thread running all the way through it. The first nine members of the group all give answers of the PQRV ('they look the same') type to problem R5. The remaining five members all give a lot of PQRE responses, making them like group 3, as do some of the first nine members. The group thus has certain similarities to group 3.
Group 3 (16 members) and group 7 (14 members) have 9 members in common.

Group 8: All the subjects in this group often solve the problems successfully by comparing the fraction of the surface of each wheel covered by the target colour. The fractions appear to be recognised rather than calculated. The group has 9 members.

Group 9: A group of subjects who make mainly fail choices accompanied by reasons based on previous choices or outcomes, and sometimes make pass choices with similar reasons. The group has 6 members.

Group 10: A group of subjects who make erratic choices and don't give any reasons for their choices other than repetitions of the experimental instructions or unqualified assertions as to their correctness. The group has 5 members.

Group 11: A group of subjects who make mainly pass choices accompanied by a reason based on previous outcomes, and sometimes make fail choices with similar reasons. The group has 7 members.

Group 12: A similar group to group 1, but group 1 is made up of PQTE and FQTE categories, whereas group 12 is biased toward PQTE, especially in response to problems R2 green, R3, and R4 red. Group 1 (23 members) and group 12 (22 members) have 18 members in common.

Group 13: An almost identical group to group 3.
Group 3 (16 members) and group 13 (16 members) have 15 members in common.

Group 14: Almost identical to group 12, and similar to group 1 in the same way as group 12 is.
Group 1 (23 members) and group 14 (20 members) have 16 members in common.
Group 12 (22 members) and group 14 have 19 members in common.

Group 15: A group in which the members all give some answers involving a pass choice with comparison of the sizes of the target segments on each wheel, together with fail choices of this type, pass choices involving comparison of the fraction of each wheel covered by the target segment, or pass choices involving comparison of the relative sizes of the target and non-target segments on each wheel. The group has some similarities to group 1, group 12, and group 14.
Group 1 (23 members) and group 15 (19 members) have 16 members in common.
Group 12 (22 members) and group 15 have 16 members in common.
Group 14 (20 members) and group 15 have 13 members in common.

Group 16: A group of subjects who often seem to produce pass choices by comparing the sizes of the target segments on each wheel, especially in response to the part of problem R4 in which red is the target colour. The group has similarities to group 1, group 12, group 14, and group 15.
Group 1 (23 members) and group 16 (23 members) have 15 members in common.
Group 12 (22 members) and group 16 have 20 members in common.
Group 14 (20 members) and group 16 have 17 members in common.
Group 15 (19 members) and group 16 have 13 members in common.

Group 17: A group of subjects who make mainly fail choices accompanied by reasons which seem spurious or irrelevant by adult standards. The group has certain similarities to group 2, group 4, and group 5.

Group 2 (15 members) and group 17 (18 members) have 13 members in common.
Group 4 (13 members) and group 17 have 10 members in common.
Group 5 (12 members) and group 17 have 8 members in common.

Group 18: A group of subjects who often seem to produce pass responses by comparing the sizes of the target segments on each wheel, especially in response to the part of problem R2 involving green as the target colour. The group has similarities to group 1, group 12, group 14, group 15, and group 16.
Group 1 (23 members) and group 18 (23 members) have 21 members in common.
Group 12 (22 members) and group 18 have 17 members in common.
Group 14 (20 members)and group 18 have 16 members in common.
Group 15 (19 members) and group 18 have 15 members in common.
Group 16 (23 members) and group 18 have 15 members in common.

Group 19: A heterogeneous group of subjects who sometimes make choices without giving justifications,especially in response to the part of problem R2 involving green as the target colour, but often also give pass choices with reasons based on previous outcomes or comparisons of the sizes of the target segments on each wheel. The group seems to represent a transition period between the erratic choices of stage 1 and the more systematic model 2 strategy. It has 6 members.

Group 20: This group seems to consist of subjects who make fail choices, with a mixture of irrelevant reasons and inexplicit reasons referring to relevant factors. The group has 5 members.

These groups seem to correspond to the stages outlined in the initial 'interpretative' analysis of the results of the roulette experiment. The various response styles associated with the stage 1 types of response appear in different groups, there are several groups corresponding to the stage 2 strategy, and a few groups indicating the model 3 strategy. This fits the general observation made that there seems to be a retardation of the roulette results by comparison with the beads results, leading to an increase in the number of stage 1 answers and a decrease in the model 3 strategy. The groups also show a greater heterogeneity of categories than the groups found in the beads analyses, which accords with the view that the slower developmental trend found in this situation leads to a greater degree of stage mixture. The analysis also separates the two principal quantification methods of size estimation and fraction recognition.

On the basis of this interpretation the following clusters of groups might be anticipated:

Cluster A: A cluster of a group of subjects who choose erratically without justifying their choices (group 10), and groups of subjects who make erratic choices based on previous outcomes (group 9 and group 11), with a linking group of subjects exhibiting both forms of response (group 19). This cluster will consist of group 9, group 10, group 11, and group 19.

Cluster B: A cluster of groups of subjects who make erratic choices and give irrelevant reasons (group 2, group 4, group 5, group 17), or who make erratic choices and give a mixture of irrelevant and relevant but inexplicit reasons (group 20). This cluster will consist of group 2, group 4, group 5, group 17, and group 20.

Cluster C: A cluster of groups of subjects who consistently choose the wheel with the lzrger sized target segment (group 1, group 6, group 12, group 14, group 16, group 18), a group of subjects who consistently choose the wheel with the larger looking fraction of target colour on its surface (group 8), a transition group of subjects mixing this strategy with a more advanced strategy (group 15), and a transition group of subjects mixing this strategy with less advanced methods of responding (group 19). This cluster will consist of group 1, group 6, group 8, group 12, group 14, group 15, group 16, group 18, and group 19.

Cluster D: A cluster of groups of subjects who consistently compare the relative sizes of the target and non-target segments on each wheel in oder to make a choice (group 3, group 13), a group of subjects who do this and yet may rely on what things 'look' like in problem R5 (group 7), and a group of subjects mixing this strategy with a less advanced strategy (group 15). This cluster will consist of group 3, group 7, group 13, and group 15.

It can be seen that of these predicted clusters, A and B correspond to stage 1, C is stage 2, and D is a third stage. In each case the transitional groups are tidied up by assigning them to the adjacent stage, as this was found to be the effect of a statistical association criterion in the analyses of the beads experiment results. It might be doubted if Cluster D will actually be separated, or if it will remain as part of cluster C, since the model 3 strategy never becomes dominant in the experimental results.

The actual clusters generated by the program were as follows:

First cluster:	Group 9 Group 11 Group 10 Group 19
Second cluster:	Group 2 Group 4 Group 5 Group 20 Group 17
Third cluster:	Group 12 Group 14 Group 6 Group 7 Group 13 Group 3 Group 8 Group 1 Group 16 Group 18 Group 15 Group 19
Fourth cluster:	Group 17 Group 19 Group 2 Group 4
Fifth cluster:	Group 15 Group 17 Group 19

(The groups in each cluster are listed in the order printed out by the program, corresponding to decreasing order of weighting to the cluster).

Thus the first cluster is the predicted cluster A and the second cluster is predicted cluster B. The third cluster combines the predicted clusters C and D, quite understandably (see above). This leaves the fourth and fifth clusters, whose appearance is quite unexpected. From the descriptions given of the first-order groups making up these clusters it is not possible to find any reason for what has happened. This means that either some kind of artifact is present in the analysis, or a way of considering the results has been overlooked.

Fortunately, the program used for the roulette analysis was more advanced than the program used in analysing the beads results, and it specified the subjects belonging to each cluster as well as the groups (see Appendix F). By inspecting the categories describing the subjects in the fourth and fifth clusters the reason for the production of these clusters can be seen. The fourth cluster consists of a group of subjects who give a variety of different responses to problems R2, R3, and R4, but who all tend to make erratic choices with irrelevant reasons on problem R5. The fifth cluster consists of a group of subjects ranged around a core of stage 2 responses based on size estimates, all of whom mix these with other types of response. This cluster is thus a cluster of subjects drawn from the hypothesised transition periods between stage 1 and stage 2 and between stage 2 and stage 3. Inspection of the initial categorisation of the roulette responses into the stages they represent (contained in Appendix E) confirms that most of the transitional subjects can be found in the fifth cluster.

<u>Discussion:</u> What exactly does this result mean? Leaving aside the fourth and fifth clusters, the other three clusters show the general pattern seen in the cluster analysis carried out on the beads experiment results for the primary school children only. The groups of the first and second clusters show no overlap with each other, whilst the groups in the first cluster overlap with those in the third cluster, and the subjects in the second cluster overlap with the subjects in the third cluster. This is consistent with a development of the following type:

$$\begin{tikzcd} \text{Stage 1A} \arrow[dr] & \\ & \text{Stage 2} \\ \text{Stage 1B} \arrow[ur] & \end{tikzcd}$$

In the analysis of the results of the first beads experiment evidence for a further development from stage 2 to a stage 3 was adduced by showing that one of the clusters contained two distinct but overlapping stages. If the same principle is applied to the current analysis the following result is obtained:

Subjects who are only in groups of the third cluster corresponding to stage 2:

SR22, 34, 37, 39, 41, 42, 45, 48, 52, 53, 55, 56, 57, 64, 67, 68, 69.

Subjects who are in groups of the third cluster corresponding to stage 2 and in groups corresponding to stage 3:

SR23, 24, 35, 36, 38, 43, 46, 50, 51, 58, 59, 66, 70, 71, 72.

Subjects who are only in groups of the third cluster corresponding to stage 3:

SR44, 47, 60.

In the above list subjects who are also in the groups of the first and second clusters (stage 1) have been left out, so that only the possible stage 2 to stage 3 development is considered. There is evidently a large transitional overlap between stage 2 and stage 3, and it can also be seen that stage 3 is only attained by a few of the subjects. It cannot be stated unequivocally whether this is due to the upper ceiling of the age range sampled being too low, or if stage 3 never attains the prominent position it occupies in the beads type of situation. However, it has been argued that the inaccuracy of size estimation will limit the convenience and efficacy of the stage 3 strategy, and that the development of fraction recognition as a method of estimation renders it unnecessary.

The presence of two distinct versions of the stage 1 type of response which was found in the analysis of the beads experiment is repeated here, and may be worth following up. At present it looks more like a difference in the subjects' response styles than a difference in the way they cope with the problems, but this too is open to dispute.

The significance of the fourth and fifth clusters is difficult to assess. They were not anticipated, but the factors responsible for them can be easily seen in the results. The problem is to decide whether these factors are important ones which have been overlooked, or ones which are understandable but of little importance. The fifth cluster, at least, seems to belong to the latter category, but the fourth cluster may be more significant.

The property common to members of the fourth cluster seems to be that they make fail choices with irrelevant reasons on problem R5. On the other problems some of the subjects in the cluster also choose erratically with various reasons given, but many go consistently for the wheel with the larger sized target segment. In other words, many of the members of this cluster give stage 2 performances on problems R2 - R4, and stage 1 on R5. This seems to imply that problem R5, in which each wheel is half red and half green, is more difficult for these subjects than the other problems. This is very interesting because the size discrimination necessary to make the stage 2 type of response is no harder in this than the other problems, but the equal proportions of red and green on each circle is very striking in this case (to adults anyway). It may well be that these subjects are sufficiently aware of this property to drop the stage 2 strategy, but unable to put anything in its place. Put simply, the cluster may show that the application of any strategy is always based on a prior consideration of its plausibility and in problem R5 the stage 2 strategy may look implausible to some subjects. If this interpretation is correct it is a useful caution against under-estimating what children can do if conditions are favourable, and above all against assuming that children following a particular strategy will limit their attention to the dimensions of the situation appropriate to that strategy.

Such an interpretation is obviously speculative, and goes a long way beyond the facts available at present. For this reason it will be set aside and a description of the results of the second cluster analysis, carried out on problems R1 - R5, will be given. If the fourth and fifth clusters found in the first analysis do not show up in the second analysis they can safely be considered artifacts of some feature of the first analysis. If they do show up then they are either more general artifacts or interesting groups which need investigating.

Second cluster analysis of the results of the roulette experiment.

The category numbering scheme and the list of category numbers for each subject used in the second cluster analysis of the roulette experiment results can be found in Appendix F. The result of the analysis will not be described in such detail as the result of the first analysis since it is to a large extent a repetition of the same result. Instead, the points of difference between the two analyses will be highlighted, and the clusters will be singled out for particular attention.

The second cluster analysis generated 26 first order groups of subjects, whereas the first cluster analysis produced 20. These first order groups are similar to the groups produced by the first cluster analysis, the extra groups being accounted for by groups of subjects with characteristic ways of answering problems R1 and R5 (which are to some extent similar). The exact constitution of many of the first order groups in the two analyses varies, but the reasons for the groupings appear to be closely comparable.

The second cluster analysis, like the first, produces five clusters of groups. The first three of these clusters are very much like the first three clusters from the first analysis, as inspection of the lists of subjects belonging to each cluster contained in Appendix F reveals. The first cluster produced by the second analysis is an enlarged version of the first cluster from the first analysis, in the sense that all subjects belonging to the first cluster from the first analysis are also members of the first cluster from the second analysis. The second cluster from the second analysis, on the other hand, is a condensed version of the second cluster from the first analysis (with the exception of one member). The third cluster from the second analysis, which has 22 members, has 17 members in common with the third cluster from the first analysis, which has 21 members.

Discussion: The fact that these three clusters are so similar in the two analyses indicates that they correspond to some general effect (if only a procedural artifact). The minor differences between the two sets of clusters are bound to occur because the criterion for their generation is statistical, so that with the two overlapping sets of data the co-varying categories will be slightly different in each case.

The fourth and fifth clusters generated by the second cluster analysis do not have the same members as the fourth and fifth clusters from the first analysis. Inspection of the response categories produced by its members reveals that the fourth cluster from the second analysis is a cluster of subjects who choose erratically without giving any justification in response to problem R1, and the fifth cluster is made up of subjects who choose erratically and give irrelevant reasons in response to problem R1. In both cases the clusters consist solely of subjects who choose erraticially on the other problems as well, so that they can be seen as corresponding to forms of the Stage 1 type of response. In fact cluster 4 is a subgroup of cluster 1 whilst cluster 5, which has 9 members, has 6 of these in common with cluster 2, which has 17 members. The production of these clusters thus contributes little to our understanding of the result of the experiment.

The Brimer technique can lead to highly redundant and sometimes spurious cluster formation (Satterly and Brimer, 1971), so that the fact that the fourth and fifth clusters from the first analysis are not repeated in the second analysis indicates that they are not of general significance and strengthens the view that the three clusters common to both analyses encapsulate the important dimensions of the responses. This is supported by the fact that the fourth and fifth clusters of the second analysis, which did not occur in the first analysis, are closely related to the first and second clusters generated in both analyses.

Summary of the result of the roulette experiment.

An attempt will now be made to summarise the conclusions which can be drawn from the result of the roulette experiment. The main aim of the experiment was to find out if the developmental model derived from the beads experiments was applicable to a different but related situation. In general, this appears to be the case. The results of the roulette experiment show a development from a first stage to a second stage which are closely comparable to the stages found in the beads experiment, even to the extent that the same two versions of the first stage are found. However, the stages observed in the roulette results seem to lag behind those found in the beads results by a few years, and the development to a third stage is not so noticeable in the roulette results. The greater degree of stage mixture observed in the roulette results than the beads results supports this more gradual developmental model.

A second aim of the experiment was to test the more general theoretical model which was developed from the results of the beads experiments. This test was made in an informal way by seeing how well the model would fit the results obtained. Obviously this is not as satisfying as the established method of making predictions, but the ease with which the model can account for the results is very convincing.

In terms of this model the results of the roulette and beads experiments can be seen as products of the same strategies for producing solutions, with a retardation in the roulette performances caused by the later development of reliable methods of estimation (Q_e) than reliable counting (Q_c). The lack of reliability of size estimation, and above all the lack of precision inherent in its size analogue symbols, can also be used to explain the fact that the third developmental stage is not as prominent in the results of the roulette experiment as in the results of the beads experiments. The development of recognition of fractions as a means of estimation in this type of situation may also be responsible for the partial eclipse of the third stage, as it provides a quantitative symbol which, if used in conjunction with the second stage strategy, will lead to solutions whose validity is limited only by the accuracy with which the different fractions can be recognised and ordered. This view, however, is dependent on the validity of the proposed template-matching model of fraction recognition, and would have to be modified if either of the iterative models proved to be correct.

It is perhaps also worth mentioning that throughout the analysis two overlapping sets of data were considered. These were the results from all the main experimental problems involved in the roulette experiment, R1 - R5, and the results from a subset of four of the five main problems, R2 - R5, which it was considered would be more directly comparable with the results of the beads experiments. In fact, little difference was found between these two sets of results, especially in the initial 'interpretative' analysis. In the two Brimer analyses slight variations were found, particularly in the second stage of the analysis which led to the generation of the clusters, and the availability of the two analyses allowed the common features to be emphasised over those features which seemed to be more closely linked to only one analysis.

Some possible criticisms and retorts.

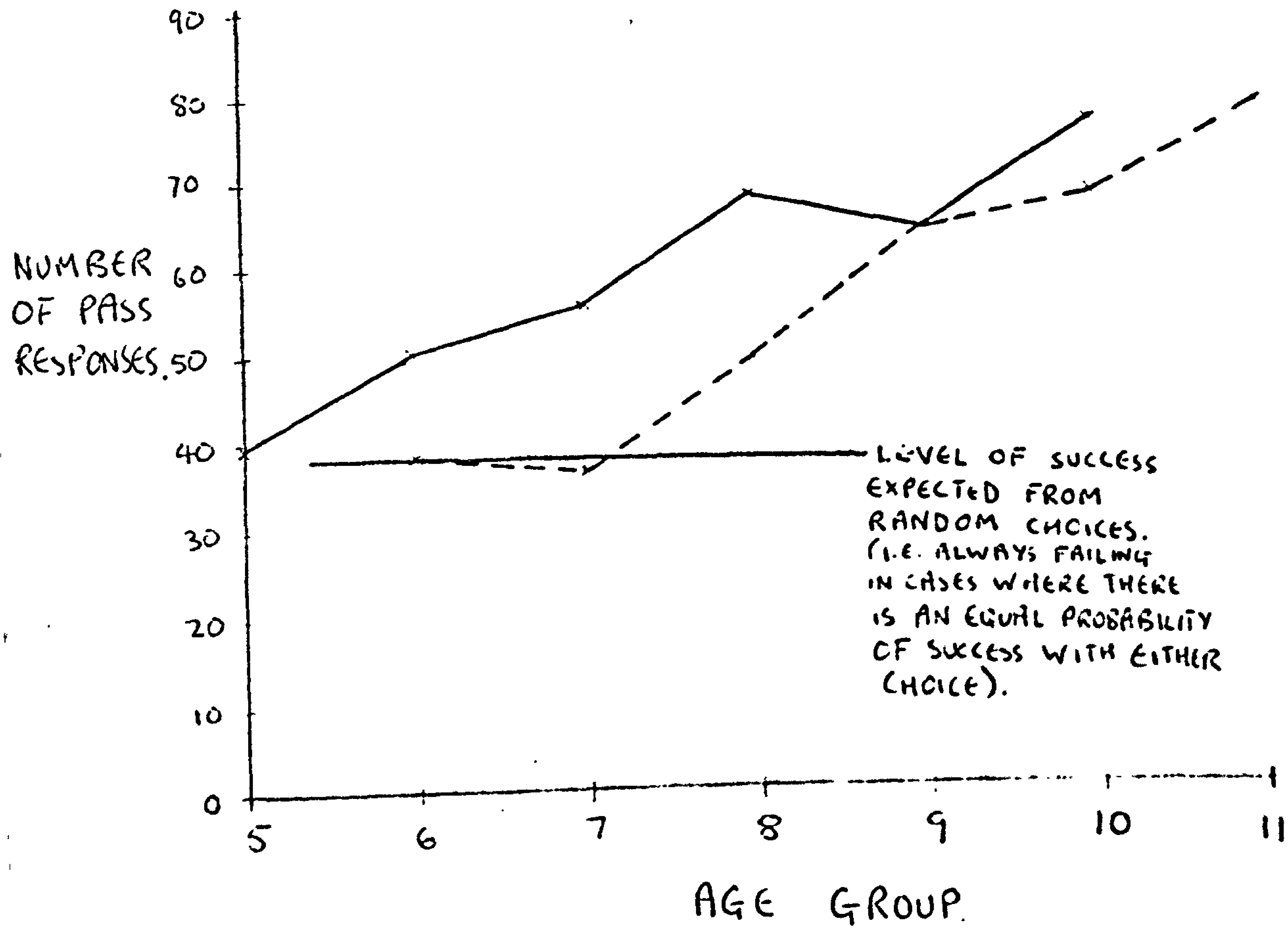
These conclusions are open to some of the same lines of attack as the conclusions derived from the beads experiments. The possibility of contamination of the results by artifacts arising from the reinforcing effect of successful gambles has again been overruled, on much the same grounds as it was ignored in the beads results, namely that this is only likely to be influencing the stage 1 subjects who don't seem to have adequate reasons for their choices.

It is possible to argue that the cause of the difference found between the results of the roulette and beads experiments may not lie in the reasons suggested, but may be due to the problems used in the roulette experiment being harder than those used in the beads experiment. For this argument 'harder' would have to mean something like 'more complicated', since the fact that the problems present more difficulty to children than the beads problems is not disputed. However, care was taken in the design of the experiment to ensure that only simple proportions of red and green were used in each problem, and in fact the roulette problems often involve simpler proportions than those used in the beads problems, so that this criticism cannot apply in any important sense.

One way in which the roulette and beads experiments did vary which might be responsible for the retardation is in manipulation of the experimental materials by the subject. In both experiments the subject made the gambles himself, but in the beads experiment he also set up the situation by putting the beads in the boxes. This may have served to draw attention to relevant aspects of the situation, but it is unlikely that this could account for the whole of the difference between the experiments, particularly in view of the ages of the subjects used. However, for the moment it cannot be entirely discounted. Offset against the greater degree of manipulation of materials in the beads experiment must be the fact that throughout the roulette experiment the relevant dimensions of the problem situation are clearly visible, whereas in the beads experiment only a representation of what is inside the boxes is used.

The use of verbal data is another feature of this type of experiment which is often criticised. It has been argued that this is legitimate if certain conditions are met, such as that the possible choices are arranged to disambiguate the reasons given in cases where this may be necessary. In the beads experiment this could be achieved without much difficulty, but the inaccuracy of estimation and the number of ways in which it can be applied severely limits the possibility of doing this

FIGURE 12: COMPARISON OF THE NUMBER OF PASS SCORES IN THE FIRST BEADS AND ROULETTE EXPERIMENTS AGAINST AGE.



FIRST BEADS EXPERIMENT, PROBLEMS B1-B4. = ———

ROULETTE EXPERIMENT, PROBLEMS R2-R5. = - - - -

in the roulette experiment. This means that the case for analysing verbal data from this experiment rests partly on the success of this method in the beads situation, and the high consistency that was then observed between the choices made and the reasons given by subjects.

It is possible that the result of the roulette experiment might not indicate a retardation in the subjects' performance as compared with the beads experiment, but merely increased difficulty experienced by the subjects in explaining how they made their choices. If this were the case it would be expected that if an analysis was made solely in terms of whether the choices made were of the collection mathematically most likely to succeed or the collection less likely to succeed, then the roulette and beads experiment results would appear the same. Such a comparison can be made quite easily, since all the choices have been divided into pass or fail choices for the Brimer analysis (a pass is a choice of the gamble which is mathematically most likely to succeed, or an assertion that either can be chosen when both possibilities are equally likely to lead to success; any other choice is a fail). By using the pass and fail categorisations of subjects' choices in the main problems of the first beads experiment, and problems R2 - R5 of the roulette experiment (the problems which are most directly comparable to those of the first beads experiment), a comparison of the number of pass scores for each of the age groups used in the two experiments can be made. This is figure 12.

Figure 12 shows that the performance of subjects on the roulette tasks is retarded by comparison with performance on the beads tasks in terms of 'correct' choices as well as sophisticated reasons. The shape of the two curves is similar, and a median test shows that for the six to ten year old group the median of the roulette experiment scores is significantly lower than the median of the first beads experiment scores (see appendix J). This means that the difference between the two experiments cannot be solely attributable to any greater difficulty experienced by the subjects in talking about the reasons for their choices in the roulette experiment.

Unfortunately this conclusion cannot be accepted without qualification as there is a feature of the roulette experimental procedure which will introduce an artifact into the pass-fail analysis. The procedure, as has already been mentioned, was not systematically randomised, but the order of problems, left or right position of large and small wheels, and orientation of target segments on the wheels was varied haphazardly by shuffling the experimental cards. However, the order of presentation of target colours on each problem was red first, then green, throughout the experiment. The advantage of doing this is that it makes it much easier to score the results of the experiment if the problem parts are kept in the same order. The disadvantage is that some of the younger subjects tend to alternate their choices between the large and small wheels, choosing the large one first, then the small one. The combination of alternation of choices with the alternation of target colours will determine the pass and fail scores of any subjects doing this. It is worth emphasising that only the stage 1 choices can be affected in this way, as stage 2 and stage 3 subjects do not alternate. Nevertheless, the tendency forms a possible source of artifact in the pass scores of the younger subjects.

The direction of this artifact will be as follows:

Problem R2: tends to decrease pass rate

Probelm R3: tends to decrease pass rate

Problem R4: tends to increase pass rate.

Problem R5: no effect.

In the cases where subjects tend to choose small first, then large, rather than large first, then small, this direction will, of course, be reversed. Hence the contribution of the artifact to the results will be dependent on how much the tendency to choose large then small dominates the opposing tendency to choose small then large. Inspection of the results gives the impression that this dominance is only slight, if it exists at all. However, the possibility of contamination exists, and the results of the analysis of pass and fail choices must be treated with consequent reserve.

Other research of interest in relation to the roulette experiment.

Before considering the next steps to be taken in the investigation, an attempt will be made to incorporate the results of the roulette experiment into the framework of other research. There are three main types of research which might be relevant in this respect. These are,

(a) Studies of probability - learning by children.

(b) Studies of children's estimates of quantities and proportions.

(c) Studies of children's performance in similar experiments to the roulette experiment.

The probability learning studies have not been mentioned until now. The situations they use are similar to the situations used in the experiments reported here inasmuch as subjects are offered choices which have different probabilities of leading to successful outcomes. In the probability learning experiment, however, the information required to calculate his chances of success is not made available to the subject on any one trial, but over a series of trials. This makes it difficult to pick out what is 'really' happening in a probability learning experiment. As Stevenson (1970) points out,

> 'The apparent simplicity of the probability learning task is deceptive......Since no solution yields consistent reinforcement, Ss are required to function as prediction-makers to a greater degree than in other learning problems'. (Op.cit., p.893).

Because of these difficulties, probability learning experiments have not yet been considered. However, there is one feature of the behaviour of subjects in such experiments which is similar to something noticed in the roulette experiment, namely that many subjects tend to alternate their choices.

This effect has been investigated by Weir (1964). Weir had consolidated data from a number of studies of probability learning involving three-choice problems to demonstrate that terminal levels of response, when plotted against age, give a U-shaped curve. The same effect was found with two-choice problems by Derks and Paclisanu (1967). One of the factors contributing to this effect was found by Weir to be the tendency of some subjects to follow a LMR (left, middle, right) or RML spatial pattern of responses. A plot of this tendency against age gave an inverted U-shaped function, with children aged seven to ten employing these patterns with a high frequency.

Sullivan and Ross (1970) have argued that this claim is oversimplified, on the grounds that many subjects use more than one alternation pattern. Nevertheless, the inverted-U relation seems to provide an interesting comparison with the alternations observed between some of the stage 1 responses in the roulette experiment. As in the probability learning experiments, these alternations tend to disappear as the children progress to other strategies, but they reappear in a much more sophisticated form later on. Most adults, for example, accept that the same collection cannot possibly be 'best' for both red and green. Paradoxically, the stage 1 children pursuing alternation strategies also perform 'as if' they know this, but the stage 2 children are quite capable of preferring the same collection for both colours. This reappearance of alternation at a later age does not seem to occur in the probability learning experiments, probably because the situations and the reasons for the alternations at the different ages are very different. This shows the dangers inherent in attempts to draw simple generalisations from complex performances. The safest thing to do at present is to avoid comparison of the results presented here with the results of probability learning experiments until a more adequate model of what the subject is doing in the probability learning situation is available.

The second research area which may be relevant to the present discussion concerns the development of children's abilities to make estimates of quantities and proportions. There is a certain amount of conceptual confusion in this field caused by the tendency of some authors to regard these different kinds of estimate as in some way equivalent. It has already been made clear that proportion involves a relation between quantities, so that an estimate of a proportion occupies a very different status to an estimate of a quantity. To illustrate this point it is worth considering the experiments reported by Ginsburg and Rapoport (1967).

As a method of investigating children's estimates of proportions, Ginsburg and Rapoport used the following procedure:

> 'He was shown an opaque container and told that in it were some black and white marbles. The E shook the box and, without looking, pulled out one marble at a time. He showed S each marble and required him to call out its colour. The marble was then transferred into a second opaque container. After the total contents of the container, forty marbles, were shown to S, he was instructed to make his estimates by means of a special apparatus. This was a large board, 23 inches long, 4 inches wide, and 1 inch deep. Along the length of the board were cut two grooves separated by 1 inch. One groove was filled with a line of forty white marbles, the other a line of forty black ones. Over each groove (and line of marbles) was a thin strip of wood which could be moved so as to reveal as many of the marbles in the line as desired. The S was shown how to do this. He was told, "What I want you to do is to show me how many white marbles and how many black ones you just saw me pull out of the box. If you saw me take out a lot of black ones, you make this (black) line very high; if I took out very few white ones, you make this line very short. If you saw about same number of black and white, you make both lines the same height. You try it now". ' (Ginsburg and Rapoport, 1967, p.207).

In experiments of this type Ginsburg and Rapoport report that six and eleven year old children are, in general, accurate in estimating proportions. However, all the tasks require the children to do is to make an estimate of the number of white beads that were in the box and an estimate of the number of black beads, and to match these to the display. How such an estimate might be arrived at is uncertain, since the beads are drawn out over what must be an extended period of time. Hence these estimates might be based on several possible criteria, such as total proportion, relative frequency of white to black, or an impression of the number of white and black separately. No firm conclusions about estimates of proportion can then be drawn from the experiment.

A similar drawback can be found in the first experiment of a series reported by Hecox and Hagen (1971). One of the aims of Hecox and Hagen's work was to investigate 'the ability of children to form accurate and consistent estimates of proportions based on a visual display' (op.cit. p. 107). The first experiment they report involves presentation of an array of 100 dots in a 10 x 10 matrix. The dots were red or green and the stimuli varied in the proportions of red and green dots, with various positionings of the red and green dots in the matrix. The subject is given an apparatus on which he can alter the proportion of red and green dots in a 10 x 10 matrix until he thinks it matches the proportion in the display. Only whole columns of the response apparatus can be made red or green, so that exact matching to the display is impossible.

This experiment provides an intriguing opportunity to test the views advanced here. Notice that the display and response matrices both have 10 columns of dots and 10 rows. This means that all a child has to do to perform successfully is to match estimates of the numbers of dots of a particular colour on each display. He does not have to estimate proportions. Key sentences in the instructions are:

> 'The idea of this game is for you to make your picture have just as many red and black dots as there are on the screen in front of you. There will be too many dots to count so I don't want you to try. This is a guessing game.' (Hecox and Hagan, 1971, p. 110).

In this situation the present view would expect the child to use Q_e to estimate the number of dots of one colour in each display and try to obtain a match. As the number of dots of the colour to which this strategy is applied increases, accuracy will go down. This is precisely the result obtained, as can be seen from the following extract:

> 'If subjects actually based their judgements on the proportion of both exhibited colours there would be no apparent reason for the disparity between performance on the high and low proportions. One possible explanation for the difference is that subjects were attending only to the red dots. As the

number of red dots increased so did the difficulty in judging their numerosity. It may be that attention strategies are important even for such relatively simple estimation tasks.' (Hecox and Hagen, 1971, p.111).

In this extract 'high and low proportions' refers to the proportion of red to black items in the display. It seems that it is the number of red items which is most likely to be estimated, but quite why this should happen isn't clear. Possibly the subject's response apparatus was reset to all black dots after each trial. At any rate in their analysis Hecox and Hagen ignore the possibility that the same strategy could have been applied to the black dots.

The second experiment reported by Hecox and Hagen (1971) uses a display of 48 closed red and green boxes containing red or green balls. The response apparatus was a 10 x 10 matrix of red or green boxes or balls. The idea is to see how well children can adjust their estimates by showing clue slides with open green boxes always containing green balls and open red boxes containing a certain proportion of red balls. A test can then be made of the subject's estimate of the proportion of red and green boxes in the display, or his estimate of the proportion of red and green balls in the boxes in the display. This experiment really does require estimates of proportions for success as there are a different number of items in the stimulus and response displays. Again, they found that what they call 'attention focussing' is a major variable in the results, able to account for the patterns of errors across the age levels studied (children of 6, 7, and 8 years). However, it is difficult to be certain as to what this experiment demonstrates, since the task is unnecessarily complex. The conclusion of Hecox and Hagen (1971) is worth quoting since it accords with the view advanced here despite the very different methods used to arrive at it,

> '.......it cannot be assumed that since a stimulus set is varying only in some quantitative dimension, responses will vary likewise. There can be significant qualitative shifts in response strategies with quantitative shifts in stimuli'. (Op. cit., p. 121).

A more satisfactory methodology for obtaining estimates of proportions from children was used by Lowe and Ranyard (1973). They showed displays containing red and green dots and then asked subjects to shade rectangles in response booklets with as much red as the proportion of red dots on each slide. They point out that this task involves:

(a) Understanding the concept of 'proportion'.

(b) Ability to estimate proportion in a display of different numbers of elements.

(c) A conversion of this estimate into a different but equivalent representation or form. (Op. cit., p. 3).

It might be thought that this would lead to 'genuine' proportional responses from children, but by carefully designing their materials Lowe and Ranyard were able to show that many children still only based their estimates on the number of red dots in the displays, even though specifically asked to judge proportions. This accords well with the viewpoint being advanced.

Finally, it is necessary to review the previous studies of children's performances in similar experiments to the roulette experiment. These are very sparse. Piaget and Inhelder (1951) investigated children's reactions to a biased roulette wheel, but the nature of this experiment is so different to that of the roulette experiment that the results are not comparable.

A series of experiments of a similar type to the roulette experiment is reported by Hoemann and Ross (1971). They start with the following contention,

> 'Neither Piaget and Inhelder nor the later investigators have, however, been concerned with the extent to which their probabilistic experimental tasks are in fact solved by the use of probability concepts'. (Op. cit., p.222).

This is not quite the same as the view advanced here, that children's solutions to probability problems do not provide direct information about the child's conception of probability.

Hoemann and Ross then try to assess the ability of children to make accurate proportionality judgements, arguing that this is an ability required for correct probability choices. However, their method for this allows a subject to succeed simply by making a magnitude judgement, and is consequently invalid. The materials they used were pairs of circles with different proportions of black and white on the surfaces, each of which was divided into a number of equal-sized black and equal-sized white segments. The circles were all the same size. Subjects were either asked to say which circle had the most of a specified colour on it, or spinners were placed on the circles and the subjects were asked which circle they would prefer to get a specified colour.

The claim made by Hoemann and Ross is that the original condition, in which subjects had to say which circle had the most of a specified colour on it, involved 'proportionality instructions', whereas the second condition, with the spinners, involved 'probability instructions'. Not surprisingly, they found no difference between the two conditions, and concluded that,

> 'Empirical support for the supposition that only comparisons between the E-designated color were performed comes from the close correlation in Experiment 1 between results with probability as compared to proportionality instructions at each odds level'. (Hoemann and Ross, 1971, p.229).

This corresponds to what has been described here as the model 2 strategy together with some form of estimation.

Hoemann and Ross then go on to discuss a curious finding which was also reported by Piaget and Inhelder (1951). When subjects are asked to predict which colour will come up in a problem involving one collection of elements of two possible colours, their performance is inferior to their performance in situations where there are two collections and the task is to pick which collection is most likely to lead to a specified outcome. The same finding is reported by Yost et al (1962), and Goldberg (1966), but in both cases their tasks differ in so many other ways that more close comparisons would be of little interest.

The explanation put forward by Hoemann and Ross to account for the effect is derived from Piaget and Inhelder (1951):

> 'Briefly put, Piaget and Inhelder claim that to perform the single-odds task correctly the child must decompose all possible outcomes into a complete set of fractions representing each different outcome. Ratios are then constructed and compared using the favourable outcomes as the numerator and the total number of possible outcomes as the denominator of each odds ratio. Thus even where only two outcomes are possible a two-step process is required.
>
> With a double-odds task, on the other hand, no decomposition of possible outcomes is necessary as long as either the numerator or the denominator is the same in both odds ratios.' (Hoemann and Ross, 1971, p. 229).

This interpretation seems to the writer to stretch credulity. The near-impossibility of making a calculation of this type with estimation as the method of quantification has already been explained, and the strategy itself would do credit to a mathematician.

In support of their interpretation, Hoemann and Ross show that in a single collection task children's performances under their 'proportionality instructions' are superior to performances under'probability instructions'. This is taken to show that subjects in the probability instructions condition do not rely on magnitude estimation in the same way as they do in the two-collections situation, but must make use of 'probability concepts'.

A simpler interpretation would be that subjects who succeed in the single-collection-and-probability-instructions condition are still using magnitude estimation, and the fall-off in performance as compared with the two collection situation is caused either by it being more difficult in this situation to see that magnitude estimation may be used, or more difficult to compare estimates of the sizes of segments of different colours than estimates of segments of one colour only. The second of these possibilities can be ruled out in view of the subject's success in the single-collection-and-proportionality-instructions condition. However, in this condition the subjects are instructed to compare the amounts of the two colours. It therefore seems that the most likely explanation of the effect is that when left to formulate their own solution strategy subjects experience some kind of difficulty in making the decision to compare magnitude estimates of unlike colours. The explanation offered by Hoemann and Ross is a particular version of this more general explanation, but other versions are possible. For example, it may be a general strategy used by children in a variety of situations to quantify all the things that look alike and choose the biggest, and this may lead to reluctance to start comparing sets of unlike things by quantification.

Another experiment carried out by Hoemann and Ross falls easily in with this interpretation. In order to make children use probability concepts in the two-collection situation they modified their task so that the children had to choose one circle to get one colour or the other circle to get the other colour. This brought performance down to the level observed in the one-collection situation. Only probability instructions were used in this experiment.

The simplest way of testing Hoemann and Ross's interpretation against the alternative interpretation offered here would be to compare performances in the type of situation used by them,where each collection is of the same total size, with performances in situations where the collections and their elements are all of different sizes. This would be what Hoemann and Ross would call a double-odds task in which neither the numerator nor the denominator are the same in both odds ratios. At present this has not been done.

The studies of children's estimates reviewed here can easily be integrated with the theory advanced, and although the studies of children's performances in experiments like the roulette experiment are less consistent with the present position, experimental tests between the different interpretations have been suggested. However, attempts to apply a theory to extant research do not provide a strong test of the theory and for this reason another experiment will be designed.

CHAPTER 5.

The Modified Roulette Experiment.

Aims for the modified roulette experiment.

The results of the roulette experiment have been considered in terms of the solution strategies identified from the results of the beads experiments and in terms of Klahr and Wallace's (1973) exposition of quantification operators. It has been argued that the subjects' performances in the roulette experiment can be seen as arising from a combination of attempts to make use of the same strategies as were observed in the beads experiments and the imprecise nature of estimation.

One possible way in which this view could be tested would be if some other means of quantification than estimation could be made available in the type of situation used for the roulette experiment. The availability of Q_s and Q_c instead of Q_e should cause results to return to the pattern observed in the beads experiment, as subjects can then generate more reliable and precise quantitative symbols.

Such a situation can be produced by drawing diameters on the wheels used in the experiment in such a way that although there is still only one red and green segment on each wheel, each of these segments is subdivided into a number of equal pieces (see the examples given in Table 6). This allows the application of Q_c in counting the number of pieces, although it will usually be invalid by adult standards.

The predictions which can then be made are:

(a) That the availability of Q_c will increase the number of stage 2 responses by the younger subjects, who are now provided with a more precise and reliable method of quantitative comparison than Q_e.

(b) That the availability of Q_c will increase the overall number of model 3 responses for the same reasons.

In both predictions the increases referred to are with respect to the result of the original roulette experiment. The predictions assume that subjects will not realise that Q_c is invalid and consequently not apply it.

The modified roulette experiment.

To test these predictions it is necessary to design an experiment which will distinguish as far as possible the forms of the model 2 strategy in which counting and estimation are the preferred means of quantification, the model 3 strategy, and the true proportion strategy corresponding to the hypothesised model 4. A modified version of the roulette experiment (which will be referred to as the modified roulette experiment) was designed for this purpose. The materials used in this experiment were similar to those used in the roulette experiment, except that the red and green segments were marked into separate pieces. This meant that the shading used to show the red and green segments had to be abandoned (it would have interfered with the lines marking the pieces) and was replaced by a system of haphazardly placed dots with an even density.

Problems used: The problems designed for the modified roulette experiment were as follows:

Problem MR1: Large wheel: 3G3R (i.e. half green and half red, each divided into three pieces).
Small wheel: 4G4R (i.e. half green and half red, each divided into four pieces).

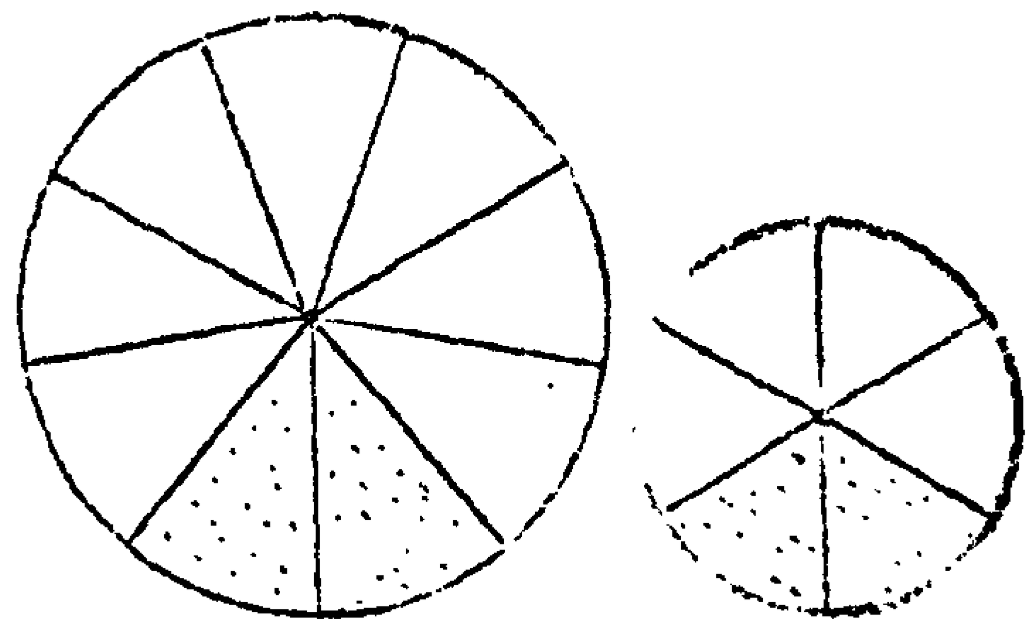
In this problem the red and green segments are largest on the large wheel, so that the use of the model 2 strategy with size estimation will lead to choice of the large wheel whether red or green is the target colour. The small wheel has more red and green segments than the large wheel, which means that the model 2 strategy together with counting of pieces will lead to choice of the small wheel in both cases. There are the same numbers of green and red pieces on each wheel, so that the model 3 strategy with counting will lead to choice of either wheel. There are also the same proportions of red and green on each wheel,so that the model 4 strategy with accurate quantification will lead to choice of either wheel. In this as in the other problems the model 2 strategy with recognition of fractions will always lead to 'correct' answers if the fractions are recognised and ordered correctly, but will lead to idiosyncratic answers if they are not recognised correctly.

Problem MR2: Large wheel: 4G2R.
Small wheel: 6G3R.

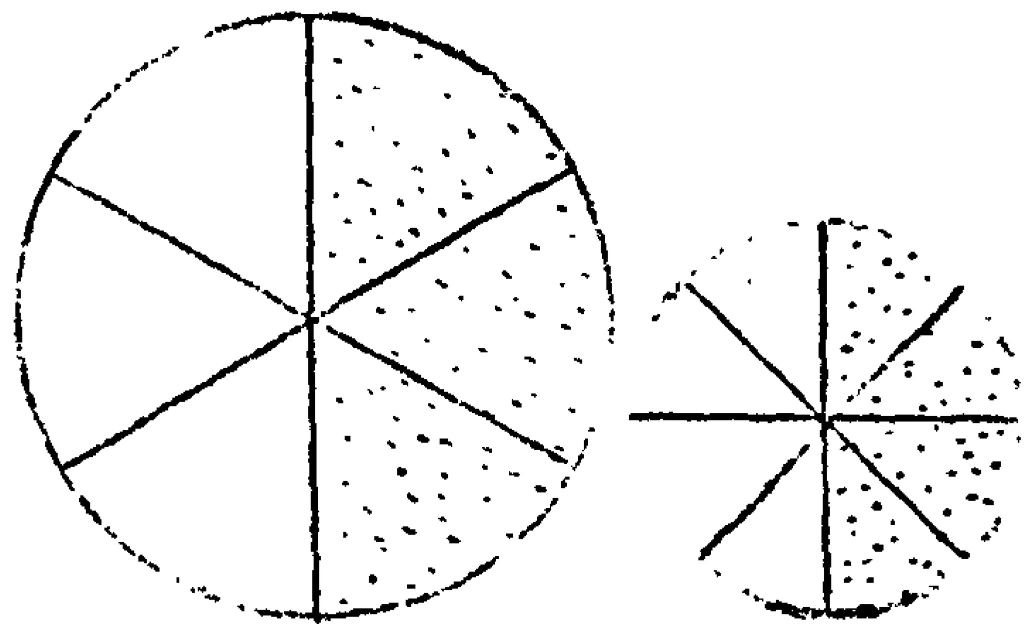
In this problem the red and green segments are again larger on the large wheel, whilst there are more red and green pieces on the small wheel. Hence the model 2 strategy will lead to choice of the large wheel when size estimation is the means of quantification and the small wheel when counting of pieces is used. The difference between the numbers of green and red pieces on each wheel is greatest for the small wheel, so that when the model 3 strategy is used with counting the small wheel will be chosen if green is the target colour, and the large wheel if red is the target colour. Both wheels have the same proportions of red and green.

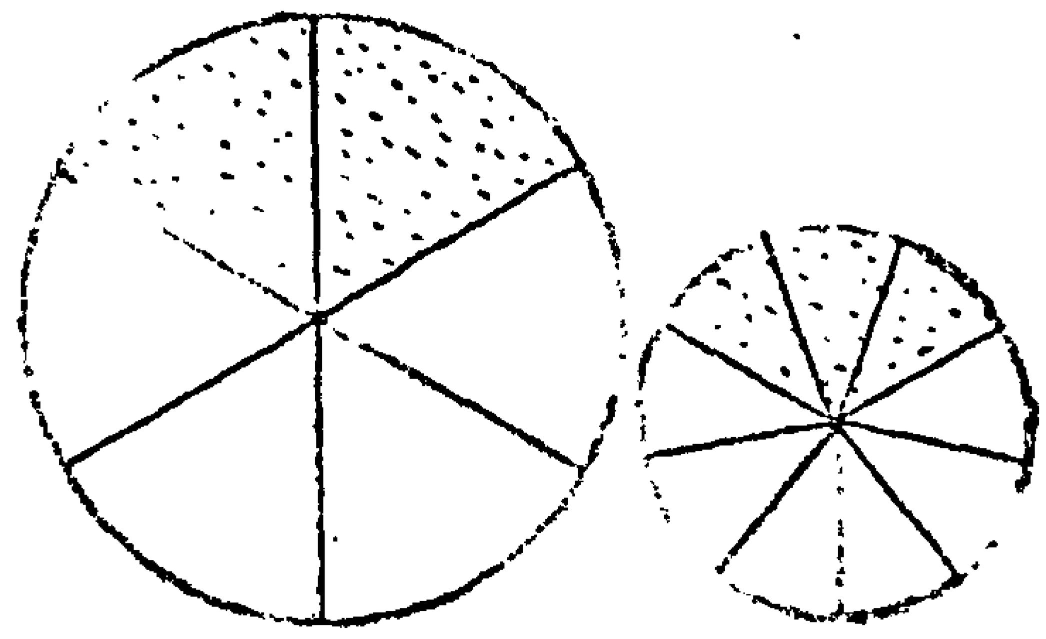
TABLE 6: PROBLEMS USED IN THE MODIFIED ROULETTE EXPERIMENT.

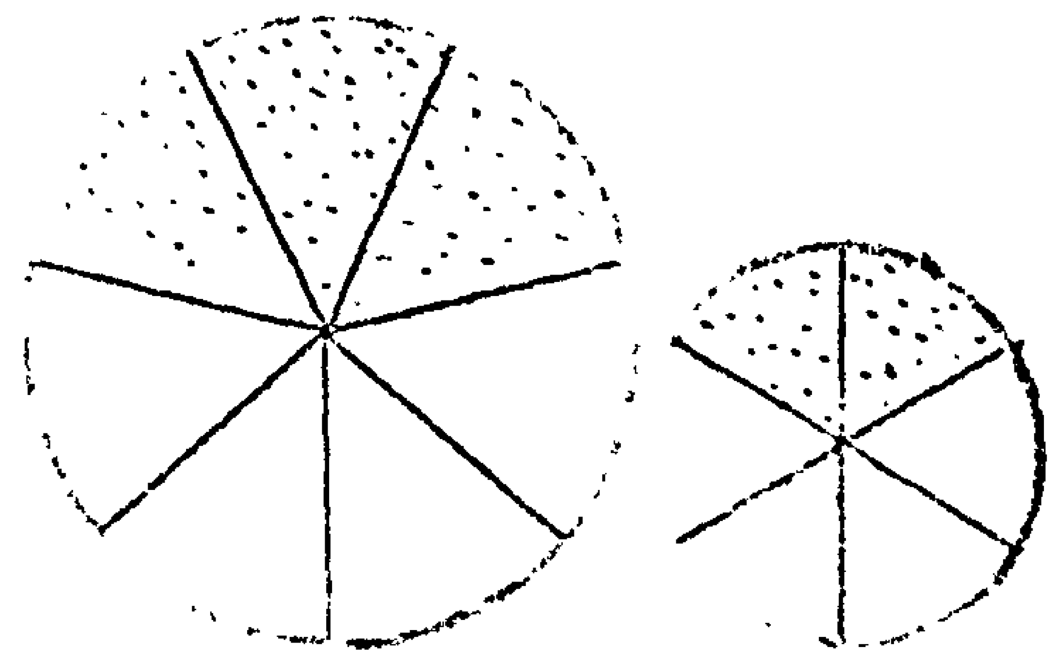
PRETEST:

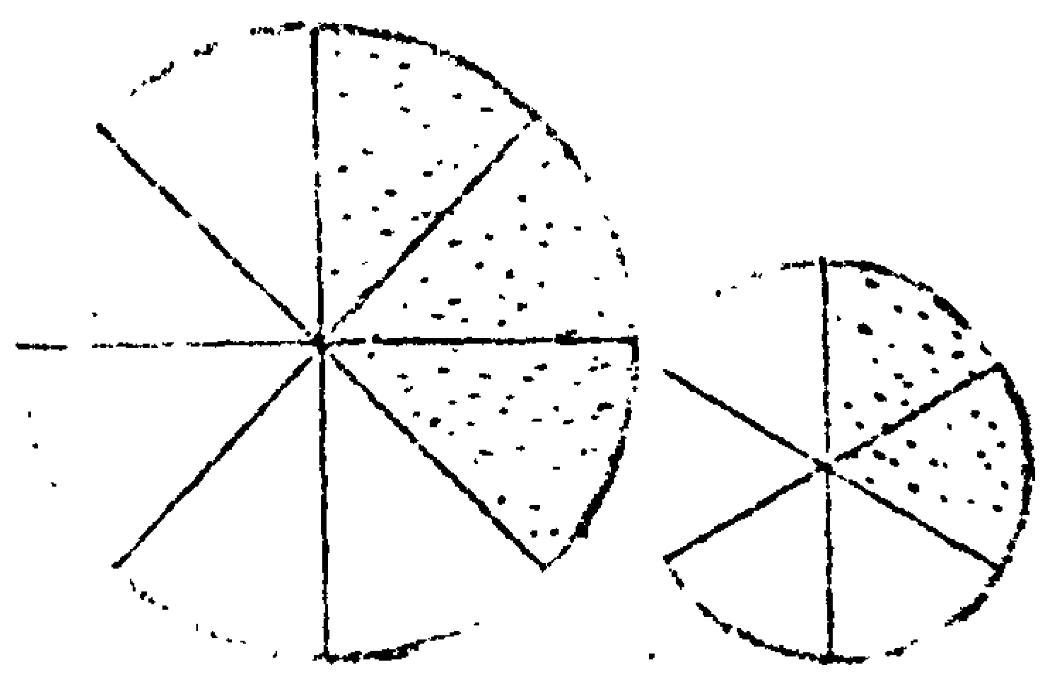
7G2RL v 4G2RS i.e. 

STUDY PROBLEMS:

MR1: 3G3RL v 4G4RS. i.e. 

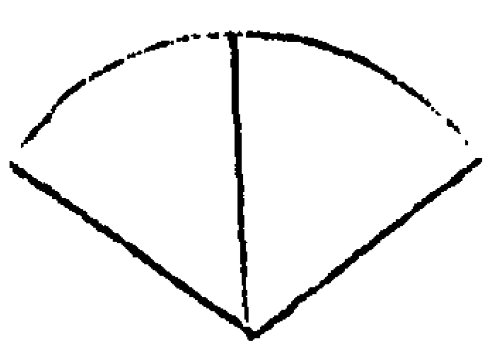
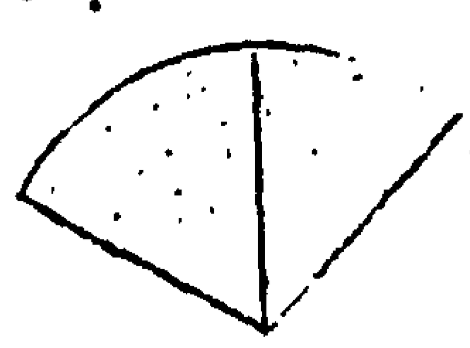
MR2: 4G2RL v 6G3RS. i.e. 

MR3: 4G3RL v 4G2RS i.e. 

MR4: 5G3RL v 4G2RS i.e. 

THE EXPERIMENT CONSISTS OF THE PRETEST, FOLLOWED BY MR1, MR2, MR3, MR4, IN RANDOMISED ORDER.

KEY: 3G3RL = LARGE WHEEL WITH THREE RED AND THREE GREEN SEGMENTS.

 = GREEN SEGMENTS.  = RED SEGMENTS.

Problem MR3: Large wheel: 4G3R.
Small wheel: 4G2R.

The red and green segments are larger on the large wheel, so that the model 2 strategy with size estimation will always lead to choice of this wheel. There are the same number of green pieces on each wheel, but more red pieces on the large wheel. This means that if the model 2 strategy is used with counting of pieces, the model 2(a) or 2(b) types of response may be observed when green is the target colour, and when red is the target colour the large wheel will be chosen. The difference between the numbers of green and red pieces on each wheel is greater for the small wheel, so that when green is the target colour use of the model 3 strategy with counting will lead to choice of the small wheel, and when red is the target colour choice of the large wheel. The most favourable proportion of green to red is on the small wheel.

In this problem the use of recognition of fractions as the method of quantification is likely to lead to idiosyncratic responses as the fractions four-sevenths and three-sevenths are not easily recognisable.

Problem MR4: Large wheel: 5G3R.
Small wheel: 4G2R.

The red and green segments are larger on the large wheel. There are also more red and green pieces on the large wheel, so that the model 2 strategy with size estimation or counting of pieces will lead to choice of the large wheel whether red or green is the target colour. The difference between the numbers of green and red pieces on each wheel is the same for each wheel, so that the model 3 strategy with counting of pieces will lead to the assertion that either wheel may be chosen in both cases. The most favourable proportion of green to red is on the small wheel.

In this problem the use of recognition of fractions as the means of quantification is again likely to lead to idiosyncratic responses as the fractions three-eighths and five-eighths are not easily recognisable.

There is also a pretest, problem MRA, which was used to allow subjects to get to know the experimental situation and requirements. This problem had 7 green and 2 red pieces on the large wheel, and 4 green and 2 red pieces on the small wheel. A list of the experimental problems, together with illustrations to assist the reader, is given in Table 6.

TABLE 6(a): SUBJECTS FROM THE MODIFIED ROULETTE EXPERIMENT WHO HAD TAKEN PART IN PREVIOUS EXPERIMENTS.

FIRST BEADS EXPERIMENT	ROULETTE EXPERIMENT	MODIFIED ROULETTE EXPERIMENT
	SR8	SMR5
	SR9	SMR11
SB7		SMR13
SB6	SR5	SMR18
SB8		SMR20
SB13		SMR23
	SR27	SMR25
	SR26	SMR26
SB24	SR16	SMR30
SB21	SR18	SMR31
	SR28	SMR33
	SR31	SMR37
	SR25	SMR39
SB33		SMR41
	SR34	SMR45
SB29	SR33	SMR46
	SR37	SMR48
SB49		SMR52
SB46		SMR55
	SR52	SMR56
	SR42	SMR57
SB47		SMR58
SB48	SR51	SMR59
	SR45	SMR60

<u>Procedure:</u> The presentation and methods used in the roulette and modified roulette experiments differed in only one respect, namely that the order of target colours, which had been alternated in the roulette experiment, was more varied in the modified roulette experiment. This was done to avoid the possible criticisms of this which were encountered in the analysis of the results of the roulette experiment. It was achieved by making the order of target colours on successive problems red-green, green-red, red-green, and so on. The combination of this with the haphazard order of presentation of the problems eliminates any significant artifact at the level of analysis to be employed. As the other methods and instructions were the same as those used in the roulette experiment they will not be repeated.

<u>Sample:</u> The subjects used were sixty children aged six to ten years, (twelve of each age) drawn from the same primary school as the children of these ages in the other main experiments. Some of the children had taken part in previous experiments, but the most recent of these had been carried out four months before. Table 6(a) lists the subjects in question. The table shows that the possibility of a practice effect must be admitted in principle. However, such an effect would militate against the direction of the observed outcome of the experiment (which is reported below), and consequently is of less importance than might have been the case.

<u>Results of the modified roulette experiment</u>

The way of treating results developed in the previous analyses will be used again for the results of the modified roulette experiment. This involves an initial analysis based on an interpretation of the strategies the subjects' responses correspond to, followed by a comparison of the result of this interpretative analysis with the analyses of the previous results, and finally an independent analysis using the Brimer technique.

The first step in the analysis thus involves specification of the criteria on which the initial interpretation of the results will be based. Essentially the same scheme as that devised for the interpretation of the results of the roulette experiment was used. This scheme will be repeated here in order to give coherence to the samples of answers of each type drawn from the experimental results, which can be compared with the samples drawn from the results of the beads and roulette experiments.

The scheme is as follows:

Type 1: Any answer corresponding to the first stage of development found in the results of the beads and roulette experiments. Such answers seem to the experimenter to involve spurious factors, or guesses. Choices made are erratic, and their justifications often seem post hoc. The following illustrations of this type of answer are taken from the results of the experiment:

SMR1 (modified roulette experiment, subject number 1), problem MR2 green (problem number MR2, target colour green).
Chooses 4G2RL (the large wheel, which has four green pieces and two red pieces marked on it). 'I dunno'.

SMR2, problem MR2, green.
Chooses 6G3RS. 'I think I'll win'.

SMR3, problem MR2, green.
Chooses 4G2RL. 'We had that one there before'.

SMR5, problem MR4, red.
Chooses 5G3RL. 'It's a bigger circle'.

SMR9, problem MR1 green.
Chooses 4G4RS. 'It just stopped on it'.

Type 2: Answers corresponding to the model 2 strategy. Such answers involve a comparison of the amount of each wheel covered by the target segment, and a choice of the wheel with the larger amount covered. The actual method used for the quantification is disregarded as long as the subject indicates that he compared the amount of each wheel covered by the target segment and chose the one with more. Answers involving comparisons of the amount of each wheel covered by the non-target segment followed by choice of the one with less are also scored as type 2, although these are rare.

Examples of type 2 answers:

SMR10, problem MR1 red.
Chooses 4G4RS. 'It's got more than that. Got four, that's got three'.

SMR10, problem MR2 red.
Chooses 6G3RS. 'Got more than that one. More red bits'.

SMR45, problem MR1 green.
Chooses 3G3RL. 'There's more green 'cos it's a bigger circle'.

SMR46, problem MR3 green.
Chooses 4G2RS. 'More green on. There's a quarter red there and a bit on this'.

SMR47, problem MR3 red.
Chooses 4G3RL. 'More red. The amount of lines tells you'.

SMR47, problem MR4 green.

Chooses 4G2RS. 'Hardly any reds 'cos it's a little circle. That other red goes round further'.

SMR47, problem MR1 green.

Chooses 3G3RL. 'It's a bigger green bit. More dots on it'.

SMR53, problem MR4 green.

Chooses 5G3RL. 'There's just one more of them lines than that one'. (This seems to be a reference to the lines separating the green pieces.)

SMR53, problem MR2 green.

Chooses 6G3RS. 'The span outwards is a bit more bigger'. (Shows with his hands and tries to measure).

SMR53, problem MR3 green.

Chooses 4G2RS. 'It's much bigger. It's more round. Looks more like a half over there, only a quarter red here.'

SMR54, problem MR4 green.

Chooses 5G3RL. 'There's five quarters. That's more. They're bigger. One more quarter than the other one'.

SMR55, problem MR2 red.

Chooses 4G2RL. 'There's two here and three there, but they're smaller'.

SMR56, problem MR1 red.

Chooses 3G3RL. 'It's bigger. The circle is. They're the same really 'cos they're both half red.'

In problem MR3, when green is the target colour, there are four green pieces on each wheel. This means that subjects using the model 2 strategy with counts of pieces will find initially that either wheel may be chosen, and the distinction between those subjects who stick to this opinion (model 2(a))and the subjects who then make use of information concerning the red pieces (model 2(b))can be introduced. The following are examples of what can happen in this situation:

SMR36, problem MR3green.

Chooses 4G2RS. 'Both got the same. I'll have the small one'.

SMR10, problem MR3 green.

Chooses 4G3RL. 'It's got more green than the other one. Maybe it's the same. I think it's more'.

SMR48, problem MR3 green.

Chooses 4G3RL. 'Got more green. Four on each but more space here'.

SMR57, problem MR3 green.

Chooses 4G2RS. 'It's got less red colours than that one. Same green colours on each'.

SMR60, problem MR3 green.

Chooses 4G2RS. 'Other has three reds, this only has two. Both got four green'.

The first of these examples corresponds to model 2(a), but was the only case of this to be found in the whole experiment. The last two examples correspond to model 2(b) with quantification by counting pieces, which was not uncommon. The other two examples show a different way of resolving the problem, which was to reapply the model 2 strategy using size estimation instead of counting. The significance of this will become clear later.

Type 3: Answers indicating use of the model 3 strategy. This involves comparison of the differences between the amounts of each wheel covered by the target and non-target segments, and choice of the most favourable case (i.e. the wheel with the largest 'surplus' or the smallest 'deficit'). As with the type 2 answer category the method of quantification is disregarded.

Examples of type 3 answers:

SMR59, problem MR1 green.

Chooses 3G3RL. 'Both the same. Both got half green and half red. Just for luck'.

SMR60, problem MR1 red.

Chooses 3G3RL, after a long pause. 'Doesn't matter which one. Both have got half red and half green'.

SMR48, problem MR1 red.

Chooses 3G3RL. 'Both the same. Three of each. It might stop on red. Could be either really.'

SMR48, problem MR1 green.

Chooses 4G4RS. 'Four of each. So I'll try it. Any will do'.

Type 4: Answers which seem to involve some understanding of proportional relationships. For example:

SMR31, problem MR1 green.

Chooses 4G4RS. 'They're both halves. It always went on green. One's just bigger than the other one, they're the same really.'

A number of responses from the results could not be fitted into this categorisation scheme very easily. These turned out to be generally of two types:

(a) Answers which are similar to the type 2 answers, but involve choice of the wheel with less of the target colour instead of the wheel with more of the target colour.

For example:

SMR10, problem MR1 green.

Chooses 3G3RL. 'It's got less than that one. Only three'.

SMR10, problem MR4 red.

Chooses 4G2RS. 'It's got less red. That's three, this is two'.

(b) Answers which are similar to the type 3 answers, but only refer to the amount of target and non-target colour on the wheel chosen, without any indication of comparison between circles. The choices made may be consistent or inconsistent with the model 3 strategy. Examples of responses of this type:

SMR21, problem MR4 green.

Chooses 4G2RS. 'It's got more green than the red'.

SMR21, problem MR3 green.

Chooses 4G3RL. 'It's got more than the red one. It's got four, red only has three'.

SMR51, problem MR2 red.

Chooses 6G3RS. 'It's got three reds. Greens have got more'.

SMR37, problem MR2 green.

Chooses 6G3RS. 'It's smaller. The green's bigger. Red is only three quarters, green is six quarters'.

Both of these types of response are unclassifiable in terms of the scheme set out so far. In the previous experiments they were rare and were usually left as unclassifiable (in the roulette experiment, answers of type (b) were included in type 3 if consistent with the strategy), but they are more common in the modified roulette experiment. The first type, (a), seems to be a mis-construction of the model 2 strategy, or an early form of this strategy, and was accordingly brought into the type 1 category. The second type, (b), looks like it is either a primitive form of the model 3 strategy, or a post hoc justification of a choice made on other grounds. Subjects giving some responses of type (b) seem to make predominantly type 1 responses to the other problems, so this type was also brought into type 1. Had most of the answers of this type been consistent with model 3 choices (as was found in the roulette results) they could possibly have been considered as type 3, but this didn't happen.

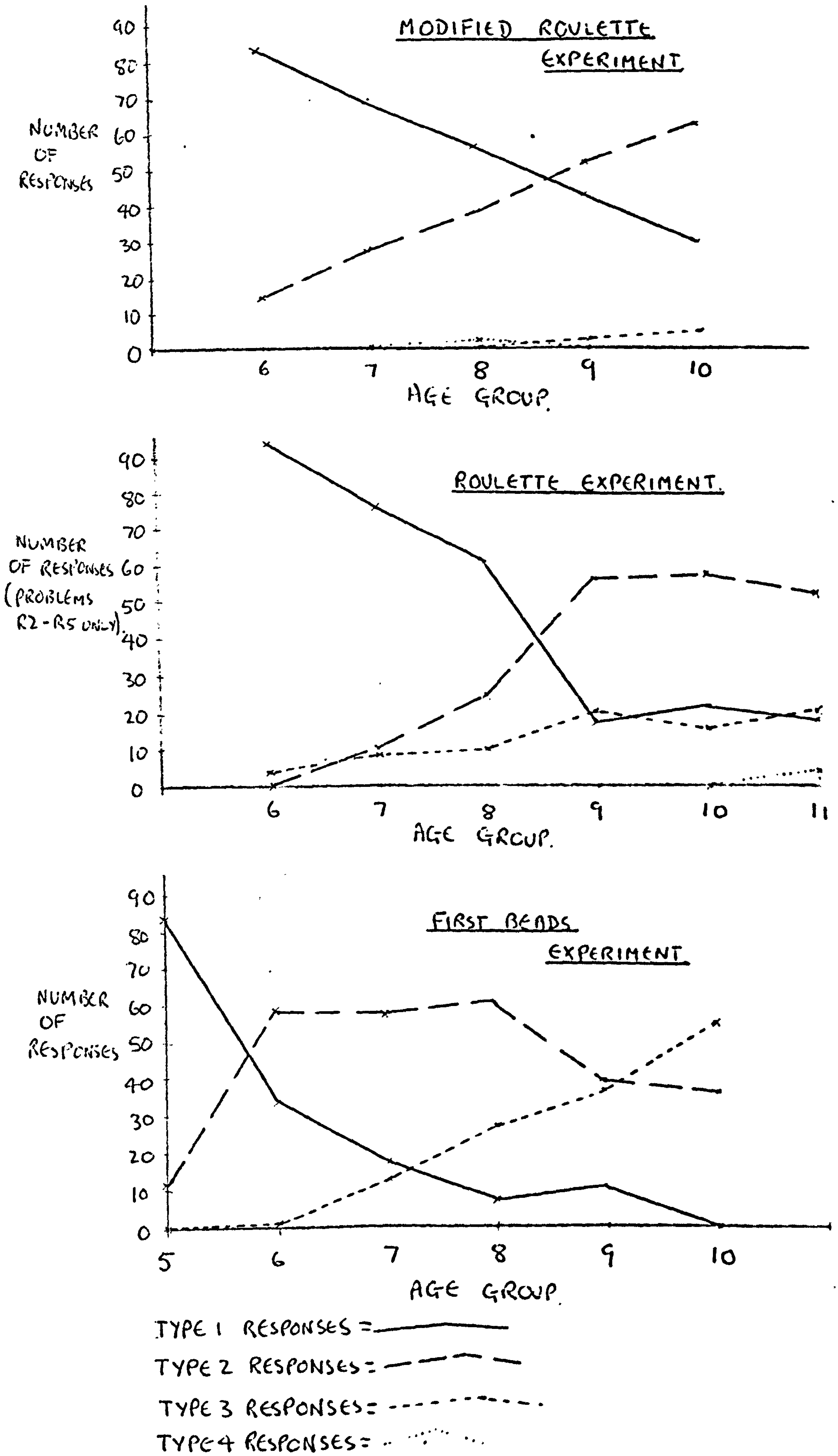
This modification to the classification scheme leaves only three unclassifiable responses. These are as follows:

SMR21, problem MR1 green.

Chooses 4G4RS. 'The green bit's bigger than the red bit. I can tell by all the dots'.

ABLE

FIGURE 13: GRAPHICAL ILLUSTRATION OF THE RESULT OF THE MODIFIED ROULETTE EXPERIMENT IN TERMS OF THE SOLUTION STRATEGIES USED BY CHILDREN OF DIFFERENT AGES, TOGETHER WITH EQUIVALENT GRAPHS OF THE RESULTS OF THE FIRST BEADS AND ROULETTE EXPERIMENTS FOR COMPARISON.



SMR34, problem MR4 red.

Chooses 4G2RS. 'There's quite a lot in each red one'.

SMR41, problem MR1 green.

Chooses 3G3RL. 'It looks bigger. The green bits look bigger than the red bits'.

The result of applying this classification scheme to the data collected in the modified roulette experiment can be found in Appendix G.

In order to facilitate comparison of the results of the modified roulette experiment with previous results, a graphical illustration can be made. This involves no particular difficulty as there are eight results for each subject and twelve subjects in each age group, and the previous graphs are also of this type. It might be though that such a comparison is unfair, as the design of the modified roulette experiment necessitated the use of two problems in which the proportions of red and green on each wheel are the same. Such problems undoubtedly skew the results of the experiment in terms of pass and fail choices, but it was shown in the analysis of the results of the roulette experiment that they do not alter the strategies the subjects use to an important degree (except in the special case where the proportions are half and half, which was used in all the experiments), so that no attempts at compensation need be made.

The graph of the result of the modified roulette experiment, in terms of the number of responses of each type for each age group, together with equivalent graphs of the results of the first beads and roulette experiments, is given in figure 13.

Inspection of this figure shows that the result of the modified roulette experiment differs from the result of the beads experiment in much the same way as the roulette experiment. That is, there is an overall retardation of the development of the different types of response, with the third type never becoming dominant. A chi-square test carried out in the same way as the comparison for the roulette and first beads experiments shows that the proportions of subjects aged six to ten years falling into stages 1, 2, and 3 in the first beads and modified roulette experiments is significantly different at the .001 level (See Appendix J).

The difference between the results of the modified roulette and roulette experiments is less pronounced. The slopes of the curves representing the type 1 responses are almost identical in each case, and the point where the type 2 response takes over from type 1 as the dominant response is also similar. However, the curve representing the type 2 response climbs more steeply initially in the modified roulette results, so that more responses of type 2 are obtained from the younger subjects in the modified roulette experiment. There is no statistically significant difference between the median numbers of type 2 answers produced by the combined six to ten year age groups in the roulette and modified roulette experiments, but the median scores of the six and seven years olds are significantly higher in the modified roulette experiment (see Appendix J). The type 3 response hardly figures at all in the results of the modified roulette experiment, and type 4 is effectively absent. The median of the numbers of type 3 answers produced by the combined six to ten year old subjects in the roulette and modified roulette experiments is significantly higher for the roulette experiment (see Appendix J).

Before the modified roulette experiment was carried out, two predictions had been made. These were that the introduction of the possibility of counting into the roulette situation would lead to a result like that of the first beads experiment, with a larger number of type 2 responses at the younger ages (first prediction) and an overall increase in the number of type 3 responses (second prediction). The second of these predictions is clearly not supported by the result of the experiment. If anything, the reverse seems to be the case, and the type 3 responses are almost eliminated. The first prediction is also only confirmed in a weak manner, and the large increase in the number of type 2 responses anticipated has not occurred.

The reason why the predictions were not supported was not initially understood. It was noted when the predictions were made that they depend upon the children's treating the counting of pieces as an appropriate means of quantification in this kind of situation. It was thought likely that this would happen, because of the greater precision and reliability of counting than estimation. However, if for some reason counting was overruled by the children, then the overall similarity of the result to the result of the roulette experiment would be explained although the drop in the number of type 3 responses would not. This explanation would not be very appealing because of its post hoc nature and the ease with which it might be applied retrospectively to the result of almost any experiment of this type. A first step in entertaining such an explanation is to consider the methods of quantification used by the children and their relative importances.

In the discussion of the results of the roulette experiment much was made of the interaction of ways of quantifying information and ways of organising quantified information to make choices. Much the same ways of quantifying information can be seen in the results of the modified roulette experiment, with the addition of ways based on the numbers of pieces in the segments. The main methods found appear to be:

(i) Counting the number of pieces in a particular segment.

(ii) Counting the number of radii separating the pieces in a segment.

(iii) Making implicit reference to numbers of pieces by using the word 'more', e.g. 'There's more red bits here than there'.

(iv) Estimating the area occupied (size) by a particular segment.

(v) Estimating the area occupied by a particular piece.

(vi) Estimating the width of a segment.

(vii) Estimating the width of a piece.

(viii) Estimating the amount of dots on a segment, without actually giving a number.

(ix) Estimating the size of the angle a segment subtends at the centre of the circle.

(x) Recognising a certain shape of segment (or range of shapes) as corresponding to a particular fraction.

This list is by no means exhaustive. The various forms of estimation, for example, were also applied to whole wheels as well as segments of the wheels. More importantly, a quite common method seems to indicate which segment of a wheel occupies more than the other segment, but nothing else. In the discussion of the pilot for the roulette experiment it was argued that this is probably a special form of size estimation, and it seems to be this which causes some of the answers which look like type 3 at first sight, but do not lead to systematic choices and only refer to one wheel. Statements of the form ' more X's and Y's' have not been considered as being the result of a separate method, but rather as the result of a calculation performed on quantitative symbols. . (The fact that no evidence is produced to support such an assumption does not matter as it is the attention paid to number which is important. This will become clear later).

In terms of the quantification operators proposed by Klahr and Wallace (1973), methods (i) and (ii) may be the result of counting (Q_c) or subitizing (Q_s), depending on the particular circumstances. All the remaining methods (including (iii)) can be seen as forms of estimation (Q_e). At present, however, the central consideration is not so much the way the quantification is carried out as what it is applied to. For this purpose it is most important that method (iii) is not lumped in with estimation.

TABLE 8: FREQUENCY OF NUMERICAL AND NON-NUMERICAL METHODS OF QUANTIFICATION REPORTED BY SUBJECTS OF DIFFERENT AGES USING THE MODEL 2 STRATEGY IN THE MODIFIED ROULETTE EXPERIMENT.

	6 YR. OLDS.	7 YR. OLDS.	8 YR. OLDS	9 YR. OLDS	10 YR. OLDS.
NUMBER OF CASES OF USE OF NUMERICAL METHODS (C AND M).	10	13	25	18	42
NUMBER OF CASES OF USE OF NON-NUMERICAL METHODS (E).	5	16	17	36	32

1578

TABLE 7 (CONT'D):

SUBJECT	AGE (YEARS)	PROBLEM MR1		PROBLEM MR2		PROBLEM MR3		PROBLEM MR4	
		GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
MR37	9								
MR38	9								
MR39	9								
MR40	9				E		E		E
MR41	9		m	m+E			E		
MR42	9	E	E				C+E		m
MR43	9			m	E	E	m	m	
MR44	9	E		E	E	E	E+C	m	C
MR45	9	E	E	E	E	E	E	E	E
MR46	9	E	E	E	E	E	E	E	E
MR47	9	E	E	E	E	m	m	E	m
MR48	9			C	C	C/E		C	C
MR49	10								
MR50	10								
MR51	10								
MR52	10			C			C		C
MR53	10	E	E	E	E	E	E	C	m
MR54	10	C/E	C	C/E	C	C/E	E	C+E	C
MR55	10	m/E	C	E	C/E	E	E	C	C+E
MR56	10	E	E	E	E	E	E	E	C
MR57	10	m	E	m/E	m	m	m	m+E	m
MR58	10	E	E	C	E+C	C	m	m	m+E
MR59	10			m	m	C	m	m+E	
MR60	10			C	C	C	C	C	C

TABLE 7: METHODS OF QUANTIFICATION REPORTED BY SUBJECTS USING THE MODEL 2 STRATEGY IN THE MODIFIED ROULETTE EXPERIMENT.

SUBJECT	AGE (YEARS)	PROBLEM MR1 GREEN	PROBLEM MR1 RED	PROBLEM MR2 GREEN	PROBLEM MR2 RED	PROBLEM MR3 GREEN	PROBLEM MR3 RED	PROBLEM MR4 GREEN	PROBLEM MR4 RED
MR1	6								
MR2	6								
MR3	6								
MR4	6								
MR5	6								
MR6	6								
MR7	6								
MR8	6								
MR9	6								
MR10	6		C		M	E		C	
MR11	6	E	E	C	E				
MR12	6	C/E	C	M	C		C	C	
MR13	7								
MR14	7								
MR15	7								
MR16	7								
MR17	7								
MR18	7								
MR19	7								
MR20	7	C				E		C	
MR21	7		E	E	C				E
MR22	7					C	C	E	E
MR23	7	E	E	E	E	E	E	E	E
MR24	7	C	M/E	C	C	C/E	C	C	C
MR25	8								
MR26	8								
MR27	8								
MR28	8								
MR29	8	E							
MR30	8			M					
MR31	8					E			
MR32	8	C			C		C	C	C
MR33	8	C		C	C		C	C	C
MR34	8	E	E	E/M	E	E	C+E	E+M	
MR35	8	E	E	E	E	E	C	C	C
MR36	8	C	C	E	C	C	E+C	M/E	C

(CONTINUED OVERLEAF).

To make this point clearer, consider what alternatives a child using estimation has available to him. If he wishes to quantify the sizes of segments he can only use the various different cues. If, however, he estimates that one segment has more pieces than another, this is something he could have confirmed independently by counting (with the numbers of pieces used in the modified roulette experiment). None of the other ways of estimating offer such a possibility. (Although it is true that in the modified roulette experiment sufficient information is available, in the form of numbers of equal sized pieces, to enable children to calculate as well as recognise fractions, no children were seen to do this. It can thus be assumed that within the age range used this kind of check is not possible).

What is needed then is to consider the results of the modified roulette experiment in terms of the difference between quantifications which involve numbers or could be brought to a numerical form, and quantifications which are not numerical in any way (including, for present purposes, recognition of fractions). This leads to the construction of table 7.

In table 7 the way in which all the type 2 responses found in the modified roulette experiment seem to have been quantified is set out. In this table the ways previously referred to as (i) and (ii) are called C, methods (iv) to (x) are referred to as E, and method (iii) is called M to show its special status as a form of estimation which alludes to counting. Where more than one form of quantification seems to have been involved this is also indicated according to the following conventions:

A/B: A quantification is carried out by a method of typeA, but is overruled by a conflicting result arrived at by a method of type B. (The quantifications are not necessarily reported as being carried out in the order AB).

A+B: Two quantifications are reported in the order given, but they both lead to the same choice and so do not conflict.
(A and B are used to refer to any of the types of quantification method).

By counting every instance of C, M, and E, for each age group a table of the frequency of quantifications by numerical methods (C and M) and non-numerical methods (E) at different ages for subjects using the model 2 strategy can be constructed. This is table 8.

The actual figures quoted in table 8 cannot, of course, be taken very seriously as they are only derived from small samples. However, they do indicate that both numerical and non-numerical methods of quantification are important in the production of type 2 responses at all of the ages involved. This allows a rejection of the view that the result of the modified roulette experiment is caused by the subjects considering counting pieces to be inappropriate.

Inspection of table 7 reveals a most interesting phenomenon. Many of the subjects using the model 2 strategy (i.e. giving type 2 responses) seem to arrive at their answers by carrying out the strategy with both numerical and non-numerical quantifications. Furthermore, most of the subjects who don't report using both types at once, do not stick consistently to one type or the other in their answers. This gives the impression that what they are doing is applying the model 2 strategy with both numerical and non-numerical quantifications. If these quantifications agree there is no problem, but otherwise one method must be preferred to the other. Possibly this will be the method leading to the larger discrepancy between wheels, possibly the last method carried out. The following illustrations of this have been drawn from the experimental results:

C + E: SMR54, problem MR4 green.
Chooses 5G3RL. 'There's five quarters. That's more. They're bigger. One more quarter than the other one'.

E + C: SMR44, problem MR3 red.
Chooses 4G3RL. 'It's a bigger red than that one. Only three lines there, four here'.

M + E: SMR58, problem MR4 red.
Chooses 5G3RL. 'Got more on than that one and they're bigger'.

E + M: SMR34, problem MR4 green.
Chooses 5G3RL. 'There's more green. It's a bigger circle and there's more triangles'.

C/E: SMR55, problem MR2 red.
Chooses 4G2RL. 'There's two here and three there, but they're smaller'.

M/E: SMR57, problem MR2 green.
Chooses 4G2RL. 'That's got more green colours, but this one's wider'.

E/M: SMR34, problem MR2 green.
Chooses 6G3RS. 'There's more green space on it. More of the triangles, although they're smaller'.

Table 7 also shows that cases where numerical methods are overruled by non-numerical methods are more common that cases where non-numerical methods are overruled by numerical methods. In addition to this the non-numerical methods involved in the cases where two quantifications are performed are invariably size or width estimation, not recognition of fractions, which only appears in cases where a single quantification is made.

In the beads experiment the majority of subjects relied on numerical quantifications and in the roulette experiment the majority of subjects relied on non-numerical quantifications. The predictions made for the results of the modified roulette experiment were based on the assumption that when given the possibility of using either numerical or non-numerical quantifications children would choose the former. The present discussion shows that this assumption was unwarranted, and that the result of the experiment can be seen as resulting from indecision by the children as to which method to employ. On this view the increase in the number of type 2 responses given by the younger subjects is seen as resulting from the availability of counting, whilst the fact that this increase does not reach the level found in the beads experiment is caused by the difficulty of reconciling counting of pieces with size estimation. The drop in the number of type 3 responses appears to be likewise caused by the problem of deciding whether to use numerical or non-numerical methods of quantification. If one method is decided on the type 3 strategy can be executed, but carrying it out with both methods and comparing the results would be a tedious and formidable task. Of the six type 3 responses found in the results two are based on counting and four on accurate recognition of fractions.

Such an interpretation must naturally be treated with caution until it has been subjected to further tests. Before ways in which this might be done are outlined, however, the rest of the analysis of the results of the modified roulette experiment will be described.

The overall result of the experiment seems to show a development from a first stage corresponding to the type 1 responses, to a second stage corresponding to the type 2 responses. This development does not involve so much stage mixture as was found in the result of the roulette experiment, chiefly because the second stage seems to be quite stable and there is only slight indication of possible development to a later third stage. This means that the result is like the beads experiment result in that the stages found seem to be stable, but the development is of the more gradual type seen in the result of the roulette experiment. However, if the interpretation of the result in terms of the discrepancy between numerical and non-numerical methods of quantification is correct, then the apparent stability of the second stage is deceptive.

Cluster analysis of results of the modified roulette experiment.

As in the analysis of the first beads and roulette experiments the Brimer cluster analysis was applied to the results in order to gain an independent and hopefully more objective appraisal of their significance. In this case only one cluster analysis was carried out, involving all the responses to the main experimental problems.

The rationale behind the cluster analysi should now be familiar, so the description of the initial categories needed to classify the results will be begun immediately. This is most easily done by considering the problems used in the order MR3, MR4, MR1, MR2.

In problem MR3 the large wheel is divided into four green and three red pieces of equal size, and the small wheel is divided into four green and two red pieces. This means that when green is the target colour a pass choice is defined as choice of the small wheel, and a fail choice is either choice of the large wheel or an assertion that it doesn't matter which wheel is chosen. The following categories of responses can be distinguished on this item:

FU: A fail choice without any reason given, or an unqualified assertion as to the correctness of the choice, or repetition of the experimenter's instruction.

e.g. SMR2: chooses 4G3RL. 'I'll win'.

PU: A pass choice accompanied by a reason like the reasons specified for FU.

e.g. SMR15: chooses 4G2RS. 'It might land on a green one'.

FSR: A fail choice accompanied by a 'spurious' reason, such as the size of the wheel chosen, the speed at which the pointer can rotate, or the segment the pointer starts on.

e.g. SMR3: chooses 4G3RL. 'It's a bigger circle'.

PSR: A pass choice accompanied by a reason like FSR.

e.g. SMR5: chooses 4G2RS. 'It's littler'.

SMR28: chooses 4G2RS. 'The point's on green'.

FPR: A fail choice with a reason based on previous choices or previous outcomes, e.g. left-right alternation, large-small alternation, win-stay lose-shift strategy.

e.g. SMR9: chooses 4G3RL. 'I just stopped before on it'.

PPR: A pass choice with a reason like FPR.

e.g. SMR1: chooses 4G2RS. 'I got red on the other one'.

FRF: A fail choice accompanied by reference to some relevant factor, without any clear reason for introducing it.

e.g. SMR6: chooses 4G3RL. 'The green's on that side and the red's on that side'.

FQACT: A fail choice accompanied by a statement of the number of pieces of the target colour on the wheel chosen.

e.g. SMR52: chooses 4G3RL. 'It's got four'.

FQAC: A fail choice accompanied by a statement of the number of pieces of both colours on the wheel chosen.

e.g. SMR21: chooses 4G3RL. 'It's got more than the red one. It's got four, red only has three'.

PQAM: A pass choice accompanied by an assertion that there are more target than non-target pieces on the wheel chosen. No reference is made to the other wheel and no numbers are given.

e.g. SMR41: chooses 4G2RS. 'Got more greens than reds. I think it'll stop on green. Never stops on red'.

This answer combines reasons of the PQAM and PPR categories, but as it is the only example of PQAM on this item it can be left as PQAM.

PQAC: A pass choice accompanied by a statement of the number of pieces of both colours on the wheel chosen.

e.g. SMR51: chooses 4G2RS. 'That's got two, the other's got four'.

PQAE: A pass choice accompanied by an assertion that the target segment is larger than the non-target segment on the wheel chosen. No reference is made to the other wheel.

e.g. SMR39: chooses 4G2RS. 'There's more green and less red on it'.

FQTE: A fail choice with a reason based on a comparison of the sizes of the target segments on each wheel.

e.g. SMR20: chooses 4G3RL. 'The easiest 'cos it's got the most green on'.

FQTDE: A fail choice with a reason indicating comparison of the two wheels by estimating the amount of dots of the target colour on each wheel.

e.g. SMR44: chooses 4G3RL. 'Bigger green bit. Got more dots than that green one'.

FQTPE: A fail choice with a reason idicating comparison of the sizes of the target pieces on each wheel.

e.g. SMR31: chooses 4G3RL. 'These green bits are bigger, those are smaller. It might land on them'.

FQTW: A fail choice with a reason based on a comparison of the widths of the target segments on each wheel.

e.g. SMR35: chooses 4G3RL. 'These green bits are bigger, those are smaller. It might land on them.'

FQTECX: A fail choice with a reason in which a comparison of the numbers of target pieces on each wheel is overruled by a comparison of the sizes of the target segments on each wheel.

e.g. SMR24: chooses 4G3RL. 'It's got more green. Four on each but it's much larger'.

FQTEPW: A fail choice with a reason based on comparison of both the sizes the target segments on each wheel and the sizes of the pieces in the target segments on each wheel.

e.g. SMR56: chooses 4G3RL. 'Large has the best chance 'cos it's bigger and with wider triangles'.

PQTM: A pass choice with a reason based on comparison of the numbers of target pieces on each wheel. The exact number is not specified, but an assertion is made that either wheel may be chosen (there are the same number of target pieces on each wheel), and the choice is the result of a guess.

e.g. SMR36: chooses 4G2RS. 'Both got the same. I'll have the small one'.

PQTF: A pass choice with a reason based on comparison of the fraction of each wheel covered by the target segment. The fractions will not necessarily be recognised accurately.

e.g. SMR53: chooses 4G2RS. 'It's much bigger. It's more round. Looks more like a half over there and only a quarter red here'.

PQNTL: A pass choice with an assertion that there are less of the non-target pieces on the wheel chosen. No numbers are given.

e.g. SMR45: chooses 4G2RS. 'There's less of reds'.

PQNTC: A pass choice with a reason based on comparison of the numbers of non-target pieces on each wheel, and choice of the wheel with less.

e.g. SMR59: chooses 4G2RS. 'Only the two parts red, the other has three parts'.

PQNTE: A pass choice with a reason based on comparison of the sizes of the non-target segments on each wheel, and choice of the wheel with the smaller non-target segment.

e.g. SMR47: chooses 4G2RS. 'Not so much red as there'.

PQNTF: A pass choice with a reason based on comparison of the fraction of each wheel covered by the non-target segment and choice of the wheel with the smaller fraction. The fractions need not necessarily be recognised accurately.

e.g. SMR46: chooses 4G2RS. 'More green on. There's a quarter red there and a bit on this'.

PQTENTC: A pass choice with a reason indicating both size estimation and counting the number of non-target pieces.

e.g. SMR22: chooses 4G2RS. 'It's smaller and there's only two red ones'.

In the part of problem MR3 involving red as the target colour a pass choice is choice of 4G3RL and a fail choice is choice of 4G2RS. The following categories must be added to the list:

PQAPE: A pass choice with an assertion that the pieces of the target segment are larger than the pieces of the non-target segment on the wheel chosen.

e.g. SMR29: chooses 4G3RL. 'The red bits look bigger than the green bits'.

FQTL: A fail choice with an assertion that there are less pieces of the target colour on the wheel chosen. No numbers are given.

e.g. SMR10: chooses 4G2RS. 'It's got less red ones'.

PQTC: A pass choice based on a comparison of the numbers of pieces of the target colour on each wheel.

e.g. SMR52: chooses 4G3RL. 'It's got three, the other only has two'.

PQTS: A pass choice based on a comparison of the numbers of pieces of the target colour on each wheel, with a statement of how many more are on the wheel chosen.

e.g. SMR35: chooses 4G3RL. 'It's got one part more than that one '.

PQTE: A pass choice based on a comparison of the sizes of the target segments on each wheel.

e.g. SMR41: chooses 4G3RL. 'Looks like a bigger red bit than the little one'.

PQTMR: A pass choice based on an assertion that there are more radii on the target segment chosen. No numbers are given.

e.g. SMR47: chooses 4G3RL. 'More red. The amount of lines tells you'.

PQTSE: A pass choice based on an attempt to compare the angles subtended by each of the target segments at the centres of the wheels.

e.g. SMR53: chooses 4G3RL. 'Bigger span than that one'.

PQTEPW: A pass choice with a reason based on comparisons of both the sizes of the target segments on each wheel and the sizes of the pieces in the target segments on each wheel.

e.g. SMR56: chooses 4G3RL. 'There's more red. The circle's bigger and the bits are bigger'.

PQTCE: A pass choice with a reason based on comparisons of both the numbers of target pieces on each wheel and the sizes of the target segments on each wheel.

e.g. SMR34: chooses 4G3RL. 'Only two ones there and three here. Bigger space as well'.

PQTEC: A pass choice with a reason based on comparisons of both the sizes of the target segments on each wheel and the numbers of target pieces on each wheel.

e.g. SMR36: chooses 4G3RL. 'It's the biggest red. That's got two, this has three'.

This completes the list of categories needed for problem MR3. In problem MR4, the large wheel is divided into five green and three red pieces, and the small wheel has four green and two red pieces. This means that when green is the target colour a pass choice is choice of the small wheel, and a fail choice is choice of the large wheel. The following new categories are introduced:

PRF: A pass choice accompanied by reference to some relevant factor, without any clear reason for introducing it.

e.g. SMR11: chooses 4G2RS. 'Got two lots of reds on'.

FQAM: A fail choice accompanied by an assertion that there are more target than non-target pieces on the wheel chosen. No reference is made to the other wheel and no numbers are given.

e.g. SMR51: chooses 5G3RL, 'It's got more green sides than the red ones'.

PQM: A pass choice following an assertion that either wheel may be chosen as there are more green pieces than red pieces on each wheel.

e.g. SMR41: chooses 4G2RS. 'Any one. Both've got more greens on each side. Try the little one'.

PQAMS: A pass choice accompanied by an assertion that there are more target than non-target pieces on the wheel chosen. No reference is made to the other wheel, but the number of extra target pieces is given.

e.g. SMR42: chooses 4G2RS. 'More green than the red parts. Four more'.

FQTM: FQTC: FQTS: FQTMR: Fail choices with reasons like PQTM: PQTC: PQTS: PQTMR.

FQTSR: A fail choice with a reason based on comparison of the numbers of radii on the target segments of each wheel. The number of extra radii on the wheel chosen is given.

e.g. SMR53: chooses 5G3RL. 'There's just one more of them lines than on that one'.

FQTMW: A Fail choice with an assertion that the wheel chosen has more pieces of the target colour and a wider target segment.

e.g. SMR57: chooses 5G3RL. 'Got more greens and wider'.

FQTEM: A fail choice with an assertion that the wheel chosen has a larger target segment and more pieces of the target colour.

e.g. SMR34: chooses 5G3RL. 'There's more green. It's a bigger circle and there's more triangles'.

FQTMXE: A fail choice with a statement that the other wheel would be best because it has more pieces of the target colour and a larger target segment.

e.g. SMR59: chooses 5G3RL. 'Hasn't got more parts, I just want it. Bigger is the best one really'.

FQTXCXE: A fail choice based on an apparent misquantification of the number of target pieces on each wheel, followed by size estimation.

e.g. SMR36: chooses 5G3RL. 'It's the biggest. Both've got the same greens. This is the biggest green bit'.

PQNTPE: A pass choice based on comparison of the sizes of the non-target pieces on each wheel and choice of the wheel with smaller pieces.

e.g. SMR22: chooses 4G2RS. 'The other has bigger red triangles'.

When red is the target colour in problem MR4 a pass choice is choice of the large wheel (5G3RL) and a fail choice is choice of the small wheel (4G2RS). This leads to the introduction of further categories:

PQAPL: A pass choice with an assertion that the wheel chosen doesn't have as many target as non-target pieces. No reference is made to the other wheel and no numbers are given.

e.g. SMR29: chooses 5G3RL. 'Hasn't got as many red triangles as green triangles.

FQTCL: A fail choice resulting from a comparison of the numbers of target pieces on each wheel and choice of the wheel with less.

e.g. SMR12: chooses 4G2RS. 'It's got two reds. I don't want so much on the paper'.

PQTSR: A pass choice with a reason based on comparison of the numbers of radii on the target segments of each circle. The number of extra radii on the wheel chosen is given.

e.g. SMR44: chooses 5G3RL. 'There's more red on it. There's more lines. One more line'.

PQTMPE: A pass choice based on comparison of both the numbers of target pieces on each wheel and the sizes of the target pieces on each wheel.

e.g. SMR58: chooses 5G3RL. 'Got more on than that one and they're bigger'.

Problems MR1 and MR2 are different to problems MR3 and MR4 in that they have equal proportions of red and green on each wheel. A pass response then involves an assertion that either wheel may be chosen, and preference for either wheel becomes a fail. The categorisation scheme can be modified to cover this situation by adding the suffixes B, for choice of the large wheel, and S, for choice of the small wheel, to the fail categories. When this has been done the following categories must also be added to the list for problem MR1 when the target colour is green.

(In problem MR1 the large wheel is divided into three green and three red pieces, while the small wheel is divided into four green and four red pieces).

FQAES: A fail choice of the small wheel with a claim that the target segment is larger than the non-target segment.

e.g. SMR21: chooses 4G4RS. 'The green bit's bigger than the red bit. I can tell by all the dots'.

FQAWB: A fail choice of the large wheel with a claim that the target pieces are larger than the non-target pieces.

e.g. SMR41: chooses 3G3RL. 'It looks bigger. The green bits look bigger than the red bits'.

FQAFB: A fail choice of the large wheel accompanied by specification of the fraction of the wheel covered by the green and red segments.

e.g. SMR37: chooses 3G3RL. 'There's half and half. It should stop easier'.

FQTCL: A fail choice of the large wheel on the grounds that it has less pieces of the target colour than the small wheel.

e.g. SMR10: chooses 3G3RL. 'It's got less than that one. Only three'.

FQTPW: A fail choice of the large wheel on the grounds that it has wider pieces of the target colour than the small wheel.

e.g. SMR35: chooses 3G3RL. 'The green bits are wider'.

FQTEL: A fail choice of the large wheel because it has less segments of the target colour and a larger target segment..

e.g. SMR55: chooses 3G3RL.. 'Got less and it's bigger red'.

FQTFDS: A fail choice with recognition of the fraction of the wheel chosen covered by the target colour, followed by an assertion that there are more dots on the target segment of the wheel. chosen.

e.g. SMR29: chooses 4G4RS. 'It's got half. Looks like it's got more green spots'.

PQTW: An assertion that it doesn't matter which wheel is chosen, with a choice based on comparison of the widths of the target segments on each wheel.

e.g. SMR58: chooouses 3G3RL. 'Got wider ones than that. Both dead heat ones really'.

PQRC: A pass response based on the fact that there are the same number of target and non-target pieces on each of the wheels.

e.g. SMR48: chooses 3G3RL. 'Both the same. Three of each. It might stop on red. Could be either really.'

PQRF: A pass response based on the fact that there is the same fraction of target and non-target colour on the surface of of each wheel.

e.g. SMR59: chooses 3G3RL. 'Both the same. Both got half green and half red. Just for luck'.

For the part of problem MR1 in which red is the target colour only one more category is needed:

FQTEMX: A fail choice based on overruling of the wheel with more pieces of the target colour in favour of the wheel with the larger target segment.

e.g. SMR24: chooses 3G3RL. 'Other has more, but this is bigger'.

In problem MR2 the large wheel is divided into four green and two red pieces, and the small wheel is divided into six green and three red pieces. Whether green or red is the target colour a pass response is defined as an assertion that choice of either wheel is equally likely to lead to success, and preference for one wheel over the other is regarded as a fail.

When green is the target colour the following new categories must be introduced:

FQAPE: A fail choice based on comparison of the sizes of the pieces on each wheel, and choice of the wheel with the larger pieces.

e.g. SMR22: chooses 4G2RL. 'It's got bigger squares. It's more easier'.

FQNTC: A fail choice based on comparison of the number of non-target pieces on each wheel and choice of the wheel with less.

e.g. SMR60: chooses 4G2RL. 'Only got two reds, three reds on the other'.

FQTSE: A fail choice based on an attempt to compare the angles subtended by each of the target segments at the centres of the wheels.

e.g. SMR53: chooses 6G3RS. 'The span outwards is a bit more bigger'. (Shows with his hands and tries to measure).

FQTMEX: A fail choice based on implicit comparison of the numbers of target pieces on each wheel and overruling their size differences. No numbers are given.

e.g. SMR34: chooses 6G3RS. 'There's more green space on it. More of the triangles, although they're smaller'.

FQTWMX: A fail choice based on comparison of the widths of the target segments on each wheel and overruling a claim that the wheel not chosen has more pieces of the target colour. No numbers are given.

e.g. SMR57: chooses 4G2RL. 'That's got more green colours, but this one's wider'.

Finally, when red is the target colour in problem MR2, one more category is needed:

FQTWCX: A fail choice based on comparison of the widths of the target segments on each wheel and overruling a comparison of the numbers of target pieces on each wheel. The numbers of non-target pieces on each wheel are quoted in support of the choice made.

e.g. SMR58: chooses 4G2RL. 'There's two wide ones instead of three little ones. Only four greens, six on the other one'.

This completes the list of categories needed to classify the results of the modified roulette experiment. Once again the point must be made that the category labels are intended as aids to the reader, and the categories themselves are formed by grouping together indistinguishable responses to the same experimental item. The complete categorisation of the results, together with data in the form used for the analysis, can be found in Appendix H.

The increase in the number of categories which must be introduced for the results of the modified roulette experiment is striking. The experiment involves 60 subjects each giving 8 responses and requires 177 categories. The comparable analyses from the beads and roulette experiments involved 8 responses from 72 subjects in each case, with 141 categories and 116 categories respectively. Of the categories from the modified roulette experiment results 86 are unique in the sense that they only have one member. This illustrates the great variability of responses in this experiment.

A description of the first-order groups of subjects generated by the cluster analysis program, in terms of the factors which seem to underly the groupings, will now be given. The reader can check the interpretations by consulting Appendix H.

Group 1: A group of subjects who choose erratically and give reasons for their choices which are spurious or irrelevant by adult standards. This group has 15 members.

Group 2: A group of subjects who seem to solve the problems by comparing the numbers of pieces of the target colour on each wheel. The categories of type PQTC and FQTC correspond to this method. The group has 8 members.

Group 3: A group of subjects who often make fail choices based on a comparison of the sizes of the target segments on each wheel. The group has 10 members.

Group 4: A group of subjects who choose erratically and don't give any reasons other than repetitions of the experimental instructions or unqualified assertions as to the correctness of the choices made. The group has 8 members.

Group 5: A group of subjects who seem to solve the problems by comparing the sizes of the target pieces on the two wheels. The categories PQTE and FQTE correspond to this method, and the group has similarities to group 3.
Group 3(10 members) and group 5 (8 members) have 6 members in common.

Group 6: A similar group to group 1.
Group 1 (12 members) and group 6 (11 members) have 10 members in common.

Group 7: A similar group 1 and group 6.
Group 1 (12 members) and group 7 (10 members) have 9 members in common.
Group 6 (11 members) and group 7 have 7 members in common.

Group 8: A similar group to group 4.
Group 4 (8 members) and group 8(8 members) have 7 members in common.

Group 9: A group of subjects who solve problem MR1 by comparing the fractions of each wheel covered by the target and non-target segments. The group has 3 members.

Group 10: A group of subjects who give answers of the PQNTC and PQTM categories to problem MR3. The group has 4 members.

Group 11: A group with points of similarity to groups 1, 6, and 7, but covering a wider range of answers. Some of the members of the group give more unclassifiable than 'spurious' reasons. The group has 8 members.

Group 12: A group like group 5 (with a lot of examples of the categories PQTE and FQTE), with some similarity to group 3 (which has a core of FQTE responses).
Group 3(10 members) and group 12(7 members) have 5 members in common.
Group 5 (8 members) and group 12 have 6 members in common.

Group 13: A group like group 5 and group 12, with some similarity to group 3.
Group 3 (10 members) and group 13(7 members) have 4 members in common.
Group 5(8 members) and group 13 have 5 members in common.
Group 12 (7 members) and group 13 have 6 members in common.

Group 14: A similar group to group 1, group 6, and group 7.
Group 1 (12 members) and group 14 (11 members) have 9 members in common.
Group 6 (11 members) and group 14 have 8 members in common.
Group 7 (10 members) and group 14 have 8 members in common.

Group 15: A group of subjects who make erratic choices and give answers which are spurious or irrelevant by adult standards, especially on problems MR1 and MR2. The group has some similarity to group 1, group 6, group 7, and group 14.
Group 1 (12 members) and group 15 (8 members) have 7 members in common.
Group 6 (11 members) and group 15 have 5 members in common.
Group 7 (10 members) and group 15 have 7 members in common.
Group 14 (11 members) and group 15 have 6 members in common.

Group 16: A similar group to group 1, group 6, group 7, and group 14.
Group 1 (12 members) and group 16 (10 members) have 8 members in common.

Group 6 (11 members) and group 16 have 9 members in common.
Group 7 (10 members) and group 16 have 7 members in common.
Group 14 (11 members) and group 16 have 9 members in common.

Group 17: A group of subjects who make erratic choices and give answers which are spurious or irrelevant by adult standards, especially in response to problems MR3 and MR4. The group has some similarity to group 1, group 6, group 7, group 14, group 16.
Group 1 (12 members) and group 17 (12 members) have 8 members in common.
Group 6 (11 members) and group 17 have 7 members in common.
Group 7 (10 members) and group 17 have 7 members in common.
Group 14 (11 members) and group 17 have 10 members in common.
Group 16 (10 members) and group 17 have 8 members in common.

Group 18: A similar group to group 1, group 6, group 7, group 14, group 16, and to some extent group 17.
Group 1 (12 members) and group 18 (10 members) have 7 members in common.
Group 6 (11 members) and group 18 have 8 members in common.
Group 7 (10 members) and group 18 have 8 members in common.
Group 14 (11 members) and group 18 have 8 members in common.
Group 16 (10 members) and group 18 have 9 members in common.
Group 17 (12 members) and group 18 have 7 members in common.

Group 19: A similar group to group 4 and group 8.
Group 4 (8 members) and group 19 (6 members) have 5 members in common.
Group 8 (8 members) and group 19 have 4 members in common.

Group 20: A heterogeneous group of five subjects who don't seem to belong together. The weightings of the members to this group are not very high.

Group 21: A similar group to group 1, group 6, group 7, group 14, group 16, group 17, and group 18.
Group 1 (12 members) and group 21 (8 members) have 7 members in common.
Group 6 (11 members) and group 21 have 6 members in common.
Group 7 (10 members) and group 21 have 5 members in common.
Group 14 (11 members) and group 21 have 6 members in common.
Group 16 (10 members) and group 21 have 5 members in common.
Group 17 (12 members) and group 21 have 5 members in common.
Group 18 (10 members) and group 21 have 4 members in common.

Group 22: A group like group 3 (which has a lot of examples of the FQTE categories). The group has some similarity to group 5, group 12, and group 13.

Group 3 (10 members) and group 22 (7 members) have 6 members in common.

Group 5 (8 members) and group 22 have 3 members in common.

Group 12 (7 members) and group 22 have 4 members in common.

Group 13 (7 members) and group 22 have 3 members in common.

Group 23: A group of subjects giving spurious or irrelevant answers to problems MR1 and MR2, like group 15.

Group 15 (8 members) and group 23 (8 members) have 6 members in common.

Group 24: A group of subjects giving little or no justification for their choices, like group 4, group 8, and group 19. However, group24 takes in a wider range of different responses than these groups.

Group 4 (8 members) and group 24 (8 members) have 5 members in common.

Group 8 (8 members) and group 24 have 6 members in common.

Group 19 (6 members) and group 24 have 4 members in common.

Group 25: A group giving some answers based on spurious or irrelevant factors, and some answers citing relevant factors without any clear indication as to why they were introduced. The group has 8 members.

Group 26: A similar group to group 24.

Group 24 (8 members) and group 26 (7 members) have 6 members in common.

Group 27: A similar group to group 2.

Group 2 (8 members) and group 27 (8 members) have 6 members in common.

Group 28: A group of subjects who make fail choices and give little or no justifications in response to problem MR4. The group has some similarity to group 4, group 8, and group 19.

Group 4 (8 members) and group 28 (6 members) have 5 members in common.

Group 8 (8 members) and group 28 have 4 members in common.

Group 19 (6 members) and group 28 have 4 members in common.

Group 29: A group of subjects who give same spurious or irrelevant reasons, but a variety of other reasons as well. The group has 8 members.

Group 30: A group of subjects who seem to solve the problems by comparing the sizes of the target segments on each wheel. The group has some similarity to group 3, group 5, group 12, group 13, group 22.

Group 3 (10 members) and 30 (7 members) have 4 members in common.

Group 5 (8 members) and group 30 have 4 members in common.

Group 12 (7 members) and group 30 have 4 members in common.

Group 13 (7 members) and group 30 have 3 members in common.

Group 22 (7 members) and group 30 have 4 members in common.

Group 31: A similar group to group 4, group 8, and group 19.

Group 4 (8 members) and group 31 (6 members) have 5 members in common.

Group 8 (8 members) and group 31 have 4 members in common.

Group 19 (6 members) and group 31 have 4 members in common.

Group 32: A group of subjects who give anwers of the PQTM category in response to problem MR3 when red is the target colour. The group is similar to group 10.

Group 10 (4 members) and group 32 (4 members) have 3 members in common.

Group 33: A heterogeneous group of subjects who all give at least one response with a reason based on a previous outcome or previous choice. The group has 4 members.

Group 34: A group of subjects who give a number of reasons referring to previous outcomes or previous choices. The group is similar to group 33.

Group 33 (4 members) and group 34 (2 members) have 2 members in common.

Group 35: Two subjects who give responses of category PQTDE to problem MR1 when green is the target colour, and category PQTMR to problem MR4 when red is the target colour.

One of the most noticeable things about these groups is the large amount of overlapping between them. This is due to the large number of small categories in the original results. Most of the groups seem to correspond to the strategies associated with the stages tentatively outlined. In addition, the different response styles associated with the first stages which were found in the beads and roulette cluster analyses seem to show up again. The stage 2 strategy appears in different groups in the form using counting of pieces as the means of quantification and in the form using size estimation. Recognition of fractions does not show up as a means of quantification permeating a whole group except in the special case of group 9, where it is associated with the model 3 strategy. The use of the word 'more' instead of numbers, which probably indicates the use of estimation to avoid the extra effort involved in counting, is found in groups 10 and 32.

The second stage of the cluster analysis involves the formation of clusters of the original groupings. Following the interpretation of the original groups advanced, the following clusters would be expected:

Cluster A: A cluster of groups of subjects who choose erratically and offer no justification of their choices, or a justification based on repetition of the experimental instructions, or an unqualified assertion that the choice is correct. This cluster will consist of group 4, group 8, group 19, group 24, group 26, group 28, and group 31.

Cluster B: A cluster of groups of subjects who make erratic choices and give reasons which are spurious or irrelevant by adult standards. This cluster will consist of group 1, group 6, group 7, group 11, group 14, group 15, group 16, group 17, group 18, group 21, group 23, group 25, and group 29.

Cluster C: A cluster of groups of subjects who prefer to make their choices by comparing the numbers of pieces of the target colour on each wheel and choosing the wheel which has more. This cluster will consist of group 2, and group 27.

Cluster D: A cluster of groups of subjects who prefer to make their choices by comparing the sizes of the target segments on each wheel and choosing the wheel with the larger target segment. This cluster will consist of group 3, group 5, group 12, group 13, group 22, and group 30.

However, there are other possible clusters as well, particularly in the form of combinations of these clusters. Clusters A and B may well form a single cluster of Type 1 responses, and clusters C and D overlap considerably too because of the tendency of subjects to switch from one type of quantification ot the other throughout the experiment. There may also be a transition cluster drawing together groups from clusters A or B with groups from clusters C or D, since clusters of this type were found in the beads and roulette analyses.

In addition, a number of the first-order groups do not seem to fit neatly into this pattern. These are group 9, group 10, group 20, group 32, group 33, and group 34. Two pairs of these groups seem to from the following small clusters:

Group 10 and group 32: Two groups of subjects who solve some of the problems by choosing the wheel with more pieces of the target colour without seeming to count. These groups may belong with predicted Cluster C.

Group 33 and group 34: Two groups of subjects who sometimes give reasons referring to previous choices or previous outcomes. These groups may belong with predicted cluster B.

The remaining groups are:

Group 9: The group of subjects who give type 3 answers to problem MR1. This will probably fit in with cluster D, as the other answers given by these subjects are mainly of type 2 with estimation as the method of quantification.

Group 20: This is a very heterogeneous group, and it might go anywhere.

Group 35: There are only two subjects in this group, who give mainly type 2 reasons with estimation as the method of quantification (apart from the rather unusual responses which have caused them to be grouped together). They will probably be associated with predicted cluster D.

In fact the cluster analysis program generates a surprisingly large number of clusters (seventeen, to be precise). These will be presented in the order generated, and discussed in terms of their constituent groups and the subjects listed as belonging to them. The lists of groups and subjects belonging to each cluster can be found in Appendix H, which allows the reader to form his own opinion of the description advanced.

Cluster 1: The groups making up the cluster consist of all the groups in the predicted cluster D, together with group 35. In other words, this is a cluster of subjects who often solve the problems by comparing the sizes of the target segments on each wheel, and chosing the wheel with the larger target segment. This is confirmed by inspection of the subjects listed as belonging to the cluster, as this list consists of all the subjects who give more than one response of this type except SMR55.

Cluster 2: The groups in this cluster are the groups in the predicted clusters A and B, together with group 20. This cluster, then, brings together all the groups representing the type 1 solutions. The subjects listed for the cluster are all subjects who make erratic choices and give a mixture of irrelevant justifications and no justifications.

Cluster 3: This cluster consists of the groups in predicted cluster C, groups 9 and 10, and groups from predicted cluster B (group 1, group 11, group 14, group 17, group 21). All the subjects listed for the cluster give answers arrived at by counting the numbers of target pieces on each wheel, and choosing the wheel with more target pieces.

Cluster 4: This cluster consists of the groups in predicted cluster C, together with groups 9, 10, and 32, and groups from predicted cluster B (groups 1, 14, 17, 21). The cluster is very similar in its constituent groups to cluster 3. Only one subject is listed for the cluster, who was also listed for cluster 3.

Cluster 5: Two groups from predicted cluster A (groups 28, 31), a group from predicted cluster B (group 23), and groups 32, 33, and 34.

The subjects listed for the cluster all give some answers based on previous outcomes or previous choices.

Cluster 6: The groups making up predicted cluster A, groups from predicted cluster B (groups 1,7,15,16,18,21,23,29), and groups 20 and 33. With the exception of group 33, these groups form a selection of the groups listed for cluster 2. The subjects listed for cluster 6 also form a selection of those listed for cluster 2, and give the same mixture of answers with irrelevant justifications and answers with no justifications.

Cluster 7: Four groups from predicted cluster D (groups 5,12,13,30), and group 32. The subjects listed for the cluster all give answers arrived at by comparing the sizes of the target segments on each wheel and choosing the wheel with the largest target segment.

Cluster 8: Groups from predicted cluster B (groups 1,14,15,17,21,23), predicted cluster A (groups 28, 31), and group 32. A heterogeneous collection of subjects is listed as belonging to this cluster. Some seem to give mainly type 1 responses, and some give a mixture of type 1 and type 2.

Cluster 9: This consists of groups from predicted cluster A (groups 4, 8,19,24,26,28,31), predicted cluster B (groups 18,23,29), and groups 20, 33, and 34. As only one subject is listed for all the cluster it is difficult to see what the cluster might represent. However, it obviously has some similarity to cluster 2 and cluster 6, and the subject listed is also listed for these clusters.

Cluster 10: Predicted cluster C, a group from predicted cluster D (group 30), a group from predicted cluster B (group 17), and groups 9, 10, and 32. As only one subject is listed as belonging to this cluster it is difficult to see what it represents. It looks like it may be a cluster of subjects in the transition period between the stage where mainly type 1 repsonses are given and the stage where mainly type 2 responses are given with counting of pieces as the method of quantification.

Cluster 11: This cluster consists of all the groups in predicted cluster B except for group 18, and group 2 from predicted cluster C. The thirteen subjects listed as belonging to the cluster all give some answers based on irrelevant factors. Some of these subjects give more type 2 answers than answers of the irrelevant type (which is a version of type 1), but most of the subjects give a variety of type 1 answers, including the irrelevant type.

Cluster 12: A group from predicted cluster B (group 29), a group from predicted cluster D (group 12), and groups 34 and 35. Only two subjects are listed for this cluster, both of whom show a mixture of type 1 responses and type 2 responses with size estimation as the method of quantification. In other words, this looks like a cluster representing the transition between stage 1 (type 1 responses) and the version of stage 2 (type 2 responses) in which size estimation is the preferred method of quantification.

Cluster 13: Predicted cluster C, groups from predicted cluster D (groups 5, 13, 30), and group 32. This looks like a cluster of stage 2 subjects. The two subjects listed as belonging to the cluster give answers of type 2, with counting as the preferred method of quantification.

Cluster 14: This cluster consists of most of the groups from predicted cluster B (groups 1,6,11,14,15,16,17,21,25,29), a group from predicted cluster C (group 2), and group 9. From this list of groups, and the subjects listed as belonging to the cluster, it seems that cluster 14 brings together subjects in a transition period between a stage where they give irrelevant reasons and a stage where they give type 2 reasons, with counting as the preferred method of quantification.

Cluster 15: Two groups from predicted cluster D (groups 3, 12), a group from predicted cluster B (group 23), and group 33. The four subjects listed for this group all give answers based on comparisons of the sizes ofthe target segments on each wheel. The cluster's subjects form a subgroup of the subjects listed for cluster 7, and three of them are also listed for cluster 1.

Cluster 16: Two groups from predicted cluster C (groups 2, 27), groups from predicted cluster B (groups 1,11,14,16,17,21,25), and group 9. The cluster seems to represent a transition between a stage where irrelevant reasons are given and a stage where type 2 reasons are given, with counting of pieces as the preferred method of quantification. As such it is similar to cluster 14, and three of the four subjects listed for cluster 16 are also listed for cluster 14.

Cluster 17: Groups from predicted cluster D (groups 3,5,12,30), and a group from predicted cluster B (group 23). This list of groups, and the list of subjects belonging to the cluster, indicates that the cluster corresponds to the model 2 strategy for solving the problem (the type 2 answers) with size estimation as the preferred method of quantification. The list of subjects belonging to the cluster forms an expanded version of the list of subjects belonging to cluster 15, which is similar to cluster 17.

It must be admitted that this result is disappointing by comparison with the clusters produced in the analysis of the beads and roulette experiments. The large amount of redundancy which always occurs in the first-order groupings appears to be present in the clusters as well in this case. In addition, the design of the modified roulette experiment does not allow the analysis to be repeated on a slightly different set of data, so that independent confirmation of the 'really' important clusters could be made. Of course this might be achieved by repeating the whole experiment, but this would hardly be worthwhile in view of the effort necessary and the fact that more interesting questions can be formulated from the results as they stand.

The result of the cluster analysis is not inconsistent with the expected result. Clusters corresponding closely to the predicted clusters B, C, and D, can be found, although predicted cluster A does not appear on its own. Furthermore, clusters corresponding to combinations of predicted clusters A and B, and predicted clusters C and D, are also present. As well as this there are clusters representing transition groups between what looks like a first stage (predicted cluster A and B, or type 1 responses) and a second stage (predicted cluster C and D, or type 2 responses). However, this argument would be weak were it not supported by the fact that the two previous experiments show the same trend together with a development to a possible third stage. This third stage does not appear in any important way in the results of the modified roulette experiment, but the interaction between solution strategies and the methods of quantification subjects employ comes out most clearly.

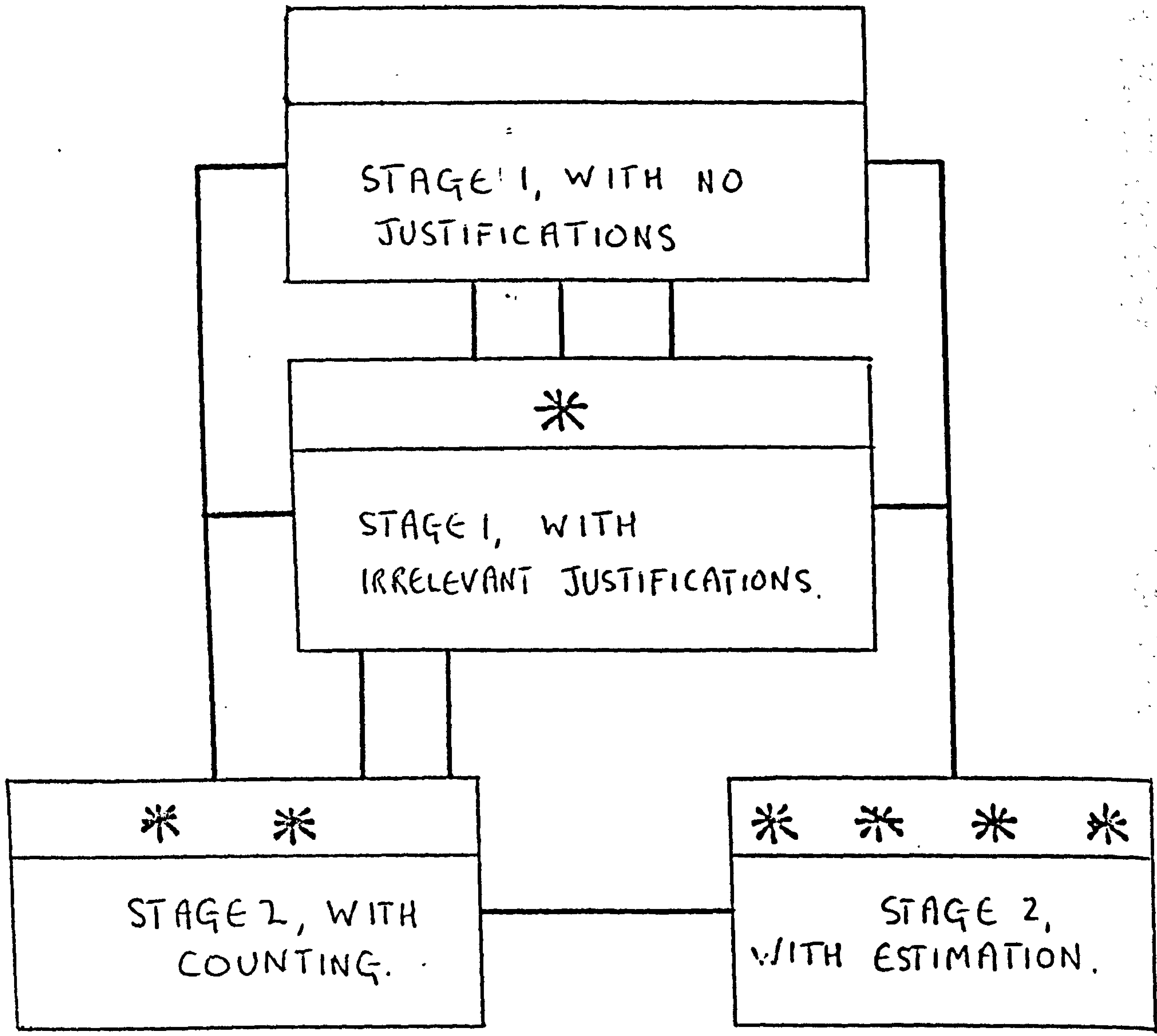
In the beads and roulette experiments the type 1 responses (or first stage of development) were divisible into the separate 'response styles' of spurious justifications or no justifications. These two different versions of stage 1 did not seem to interact with each other, but developed independently into the second stage. In the result of the modified roulette experiment the situation is very different. The two response styles still show up, but tend to be highly intermingled to the extent that only the irrelevant justifications emerge as a cluster in their own right. There are now two main versions of stage 2, corresponding to preferences for counting or size estimation as the method of quantification, and both types of stage 1 response seem to be linked developmentally to both types of stage 2 response.

This can perhaps be clarified with the aid of a diagram. Considering the clusters found in terms of the predicted clusters and their inter-relationships reveals the following situation:

Clusters corresponding to predicted cluster A (stage 1 answers of the 'no justification' variety): None.

Clusters corresponding to predicted cluster B (stage 1 answers of the 'irrelevant justification' variety): Cluster 11.

FIGURE 14: DIAGRAMMATIC REPRESENTATION OF THE 'MEANING' OF THE CLUSTERS FROM THE ANALYSIS OF THE RESULTS OF THE MODIFIED ROULETTE EXPERIMENT.



THE LINES REPRESENT LINKS BETWEEN THE TYPES OF SOLUTION WHICH SHOW UP IN THE CLUSTERS (SEE TEXT FOR EXPLANATION OF THE METHOD OF ESTABLISHING LINKS). THE ASTERISKS INDICATE THE NUMBER OF CLUSTERS CORRESPONDING ONLY TO A PARTICULAR TYPE OF SOLUTION.

Clusters corresponding to combinations of predicted clusters A and B (stage 1 answers of both varieties): Cluster 2, cluster 6, cluster 9.
Clusters corresponding to predicted cluster C (stage 2 answers with quantification of numbers of pieces): Cluster 3, cluster 4.
Clusters corresponding to predicted cluster D (stage 2 answers with estimation as the method of quantification): Cluster 1, cluster 7, cluster 15, cluster 17.
Clusters corresponding to combinations of predicted clusters C and D (stage 2 answers): Cluster 13.
Clusters corresponding to combinations of predicted clusters B and C (transition from stage 1 with irrelevant reasons to stage 2 with counting of pieces): Cluster 14, cluster 16.
Clusters corresponding to combinations of predicted clusters A, B, and C (transition from stage 1 to stage 2 with counting of pieces): Cluster 10.
Clusters corresponding to combinations of predicted clusters A,B, and D, (transition from stage 1 to stage 2 with estimation): Cluster 12.

This is represented in figure 14. Clusters 5 and 8 have been omitted from this figure as they do not seemed to be linked directly to the other clusters. Cluster 5 seems to be a cluster of subjects giving answers based on previous outcomes (probably a further form of stage 1), and cluster 8 could not be interpreted.

Summary of the result of the modified roulette experiment.

For the convenience of the reader, the conclusions which can be drawn from the analysis of the results of the modified roulette experiment will be summarised. The aim of the experiment was to test the view that the difference between the results of the beads and roulette experiments could be accounted for by the later development and general imprecision of estimation as compared with counting. To make this test an experiment was devised which, whilst remaining similar to the roulette experiment, allowed the possibility of counting as well as estimation as a method of quantification. It was predicted that this availability of counting would cause the result of the experiment to be like the result of the beads experiment rather than the result of the roulette experiment. This prediction was not confirmed.

The result of the modified roulette experiment showed a two-stage development of a gradual type, not unlike the two-stage development found in the roulette experiment. However, there was little sign of further development to a third stage. Instead, two versions of the second stage were identified, corresponding to preference for counting or size estimation as the method of quantification. There was, however, considerable intermixing of these two possibilities, and this led to a possible explanation of the result.

It is proposed that the result of the modified roulette experiment can be understood in terms of indecision on the part of the subjects as to which cue, numbers of pieces or sizes of target segments, is the one which should be quantified. This leads to much switching of cues and even some cases of comparisons of the results of quantifying both cues. This draws attention to the crucial role of the children's views as to what methods are appropriate in a given situation,which has hitherto been neglected.

The lack of any development to a third stage in the age range used for the modified roulette experiment may be explained as the result of this increased attention to different cues for quantification, instead of the persistence with one cue found in the beads and roulette experiments. The fact that recognition of fractions is not such a common form of estimation as it was in the roulette experiment is also interesting. Possible reasons for this are a reluctance by subjects to treat segments which are divided into pieces as units to be represented by a single fraction, or the fact that many of the problems used in the modified roulette experiment involve similar fractions of red and green on each wheel (this would mean that recognition of fractions would not indicate any difference between the wheels in a lot of cases, leaving the subject with the choice of asserting that either wheel may be chosen or choosing on some other basis).

Finally, the distinction between the two varieties of the stage 1 responses (those with reasons without justification and those with reasons quoting irrelevant justifications) is found in the modified roulette experiment, but the form it takes is different to that of the beads and roulette results. In these results the two types of stage 1 response were quite separate, and seemed to develop independently into stage 2, whereas in the modified roulette experiment they are overlapping and the development is complex. The reason for this is not clear, but it has been argued that the distinction is probably only one of 'response style', and may be due to the impoverished nature of the verbalisations of stage 1 children.

Some suggestions for testing the interpretation of the results of the modified roulette experiment, and some relevant research.

No general critique of the modified roulette experiment will be offered here, as thekinds of criticism anticipated have already been covered in discussions of the previous experiments. Instead, attention will be given to some ways in which the usefulness of the interpretation of the result offered might be assessed.

The first step in such an assessment must be to make sure that there is no feature of the design of the modified roulette experiment which might itself be responsible for the difference between the modified roulette and roulette experiment results. This is particularly important because the modified roulette experiment was not designed to investigate the effect which was actually found. It was designed on the assumption that counting would be used by most of the subjects, so that the materials were arranged to disambiguate answers arrived at by counting as far as possible. In fact, counting was reported by many of the subjects (otherwise the experiment would not have got past the pilot stage), but many used size estimation as well.

This meant that the lack of attention paid to the possibility of estimation in the design stages of the experiment proved problematic. In the roulette experiment only simple proportions of red and green were used, whereas in the modified roulette experiment quite complex proportions proved necessary. At first this was considered unimportant because size estimation does not depend upon proportion, and recognition of fractions is imprecise enough to lead to a 'difficult' fraction being recognised as the nearest 'easy' one known. However, the way in which the simple proportions were combined in the problems of the roulette experiment may make for 'easier' problems than those of the modified roulette experiment. Two of the roulette problems involved a wheel which was half red and half green and another wheel which was not half red and half green. In such cases it may be particularly 'easy' to see that there is the same amount of red and green on one wheel, and more of one than the other on the other wheel. This might possibly account for the drop in the number of type 3 responses in the modified roulette experiment, instead of the account advanced here in terms of excessive attention to the conflict between counting and size estimation.

A worthwhile control to test this possibility might be made, but the main claim, that the stage 2 strategy is used with counting or estimation and in some cases both, is not susceptible to such criticisms. The general lack of attention to fractions in the modified roulette experiment also could not be explained on the grounds that the fractions used were not simple, for the reasons already mentioned. This is particularly true in the modified roulette situation, where sufficient information to enable calculation of the fractions is actually provided.

A more general approach would be to investigate the criterion on which the use of numbers of pieces or sizes of segments as cues to be utilised in this type of problem is based. For example, the fact that the pieces on the two wheels in the modified roulette experiment are of different radii may favour size comparisons rather than counts. In other words, children may hesitate about using counting when the items to be compared are physically different and will then rely on estimation. Furthermore, there may be certain types of physical difference which are more important than others in this respect, such as a segment's radius rather than its angular width. The way the target pieces in the roulette experiment are kept together may also increase preference for estimation to counting, and wheels in which the target and non-target pieces are interspersed might be tried. Attention could also be drawn to the numerical aspects of the situation by allowing the subjects to 'construct' the wheels used out of red and green 'cheese pieces'. This possibility is particularly interesting as it takes up a point raised earlier about the difference between the amount of manipulation of the experimental materials by the subjects in the beads and roulette experiments.

If the criteria children use in deciding to count pieces or estimate sizes could be identified in this way, and a position could be arrived at from which the relative uses of counting or estimation by subjects (in experiments of this type) was predictable, an impressive confirmation of the interpretation advanced so far would have been achieved.

How does the interpretation of the modified roulette experiment in terms of the subjects' difficulty in deciding whether to count pieces or estimate sizes fit in with the other research literature? Quite simply, there appear to be no similar experiments (to the writer's knowledge, at any rate). It is possible, however, to find accounts of experiments where a similar conflict between cues leading to different types of quantity judgements has been set up.

The most striking of such experiments involve the investigation of what Piaget (1952) calls conservation of discontinuous quantity. These experiments involve judgements by children as to whether the relative amount of elements in two rows remains the same after transformation of one of the rows. The experiment is typically begun with both rows in one-one correspondence (same length and spacing of elements for each row), followed by a transformation in the form of lengthening or shortening of one of the rows. A non-conserving response would be to claim that initially the amount of beads in the two rows is equal and after the transformation one row has more than the other.

The explanation offered by Piaget to account for this effect amounts to a claim that the non-conserving child lacks the ability to mentally 'cancel out' the transformation. However, Bryant (1971b) argues that what actually happens depends on the fact that there are two cues which children may use to make quantity judgements in such cases, one-one correspondence and length. Initially these cues are in agreement (in the sense that they both indicate that each row has the same quantity), but after the transformation the one-one correspondence is destroyed and the length cue indicates that the longer line has more. Piaget (1952) states that the young child does not grasp that number is invariant, but Bryant (1971b) has shown that without the introduction of conflicting cues invariance is assumed, and that in the classical conservation experiment it is overruled by the conflict with the length cue. What the child must learn, then, is not that number is invariant, nor how to reverse transformations, but which quantity cues are appropriate in which situations.

The interesting thing about this, for present purposes, is the comparison between Bryant's interpretation in terms of cues for judging quantity and the analysis of the role of cues in the experiments reported here. Closer inspection of Bryant's own materials reveals that he arranges his situations carefully in order to eliminate other cues, such as counting (by using large numbers of beads). This is quite legitimate as an initial step, but it limits the scope of his theory, and a more broadly based view of the role of different types of quantitative judgements in conservation is given by Klahr and Wallace (1973).

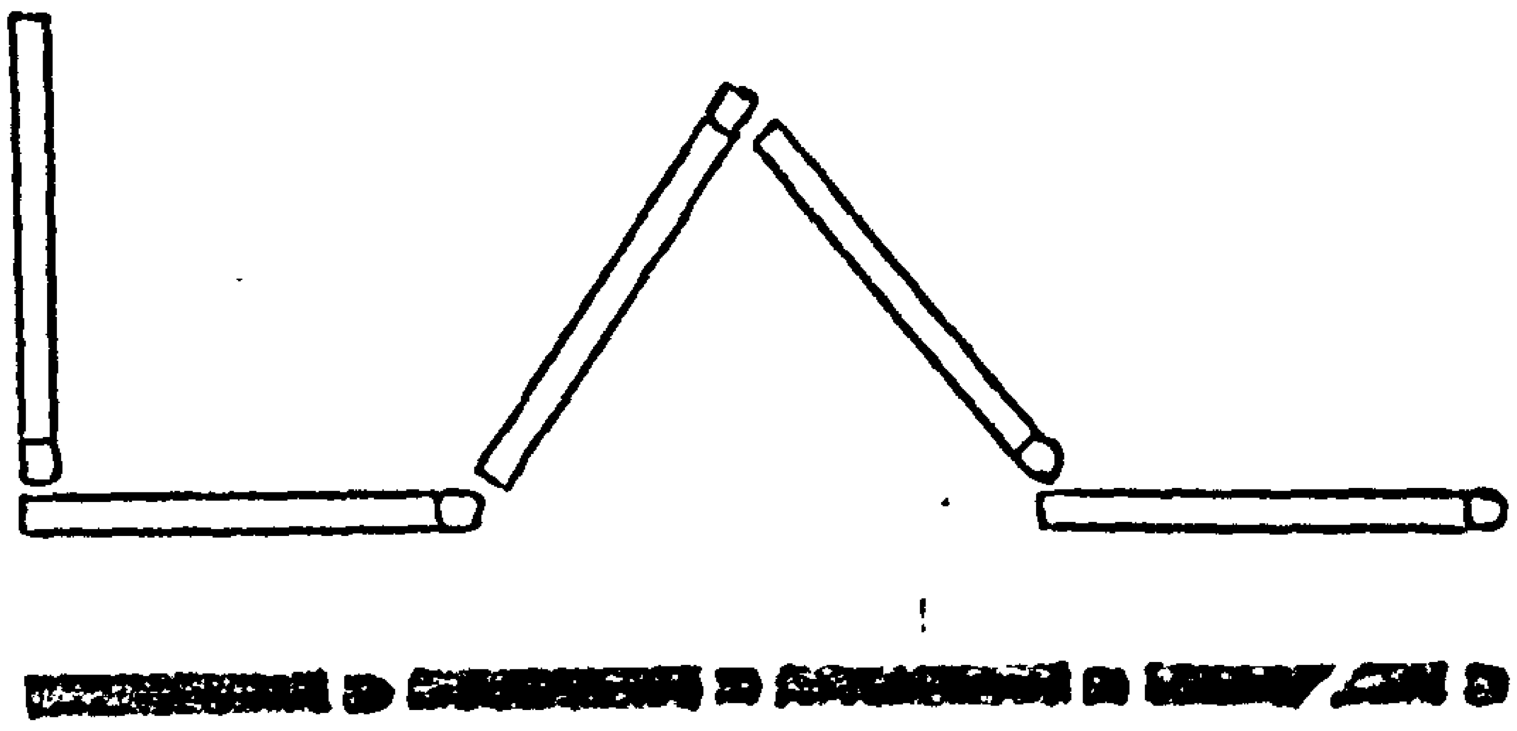
A study reported by Inhelder and Sinclair (1969) also investigates what children do when there is a conflict between different cues for assessing quantity. This study is particularly interesting for two reasons. Firstly, the cues brought into conflict are number and length, which involve Q_c and Q_c in terms of Klahr and Wallace's (1973) model, and so are comparable to the possible cues in the modified roulette experiment. Secondly, the interpretation of the results given by Inhelder (1972) is very close to the approach adopted here.

Inhelder and Sinclair's aim was to train subjects to conserve length by demonstrating that both number of constituent parts and point-to-point distance 'as the crow flies' are only valid indicators of a line's length in certain special cases. Their problems involved match sticks of two different lengths, and took three forms:

1. The experimenter constructs a zig-zag line made with one length of match, and the subject's task is to construct a straight line of the same length directly beneath, using the other length of match.
2. The subject again has to construct a straight line of the same length as the experimenter's zig-zag line using a different length of match,

FIGURE 16: SOLUTIONS OBSERVED BY INHELDER AND SINCLAIR (1969). (AFTER INHELDER, 1972, p. 109).

A.

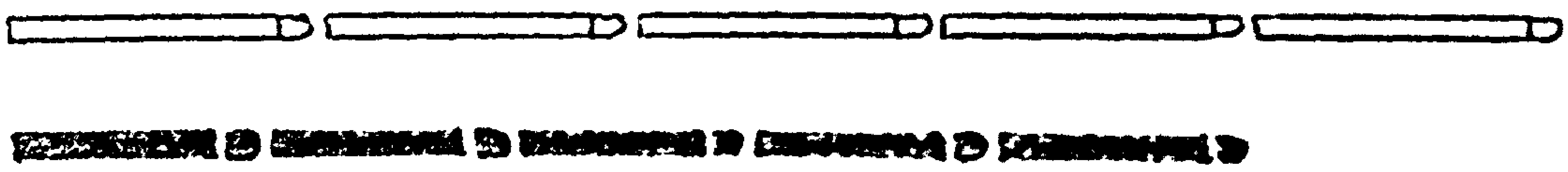


A CHILD'S ATTEMPT TO INTEGRATE ORDINAL AND NUMERICAL REFERENCES — BY BREAKING A MATCH.

B.

ANOTHER ATTEMPT TO INTEGRATE ORDINAL AND NUMERICAL REFERENCES.

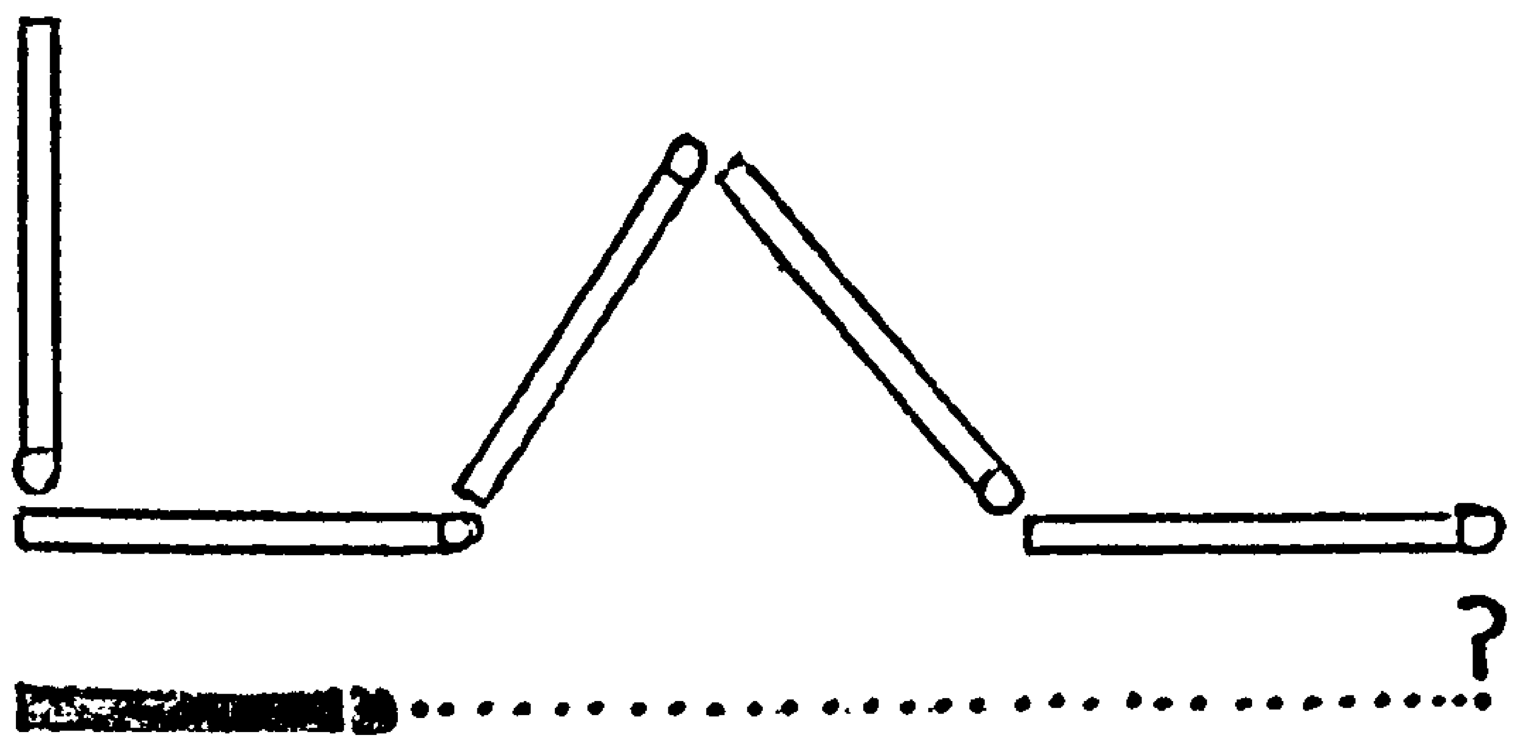
C.



IN THE THIRD PROBLEM-SITUATION, THE CHILD MAY DISCOVER THE INADEQUACY OF HIS COUNTING SOLUTION.

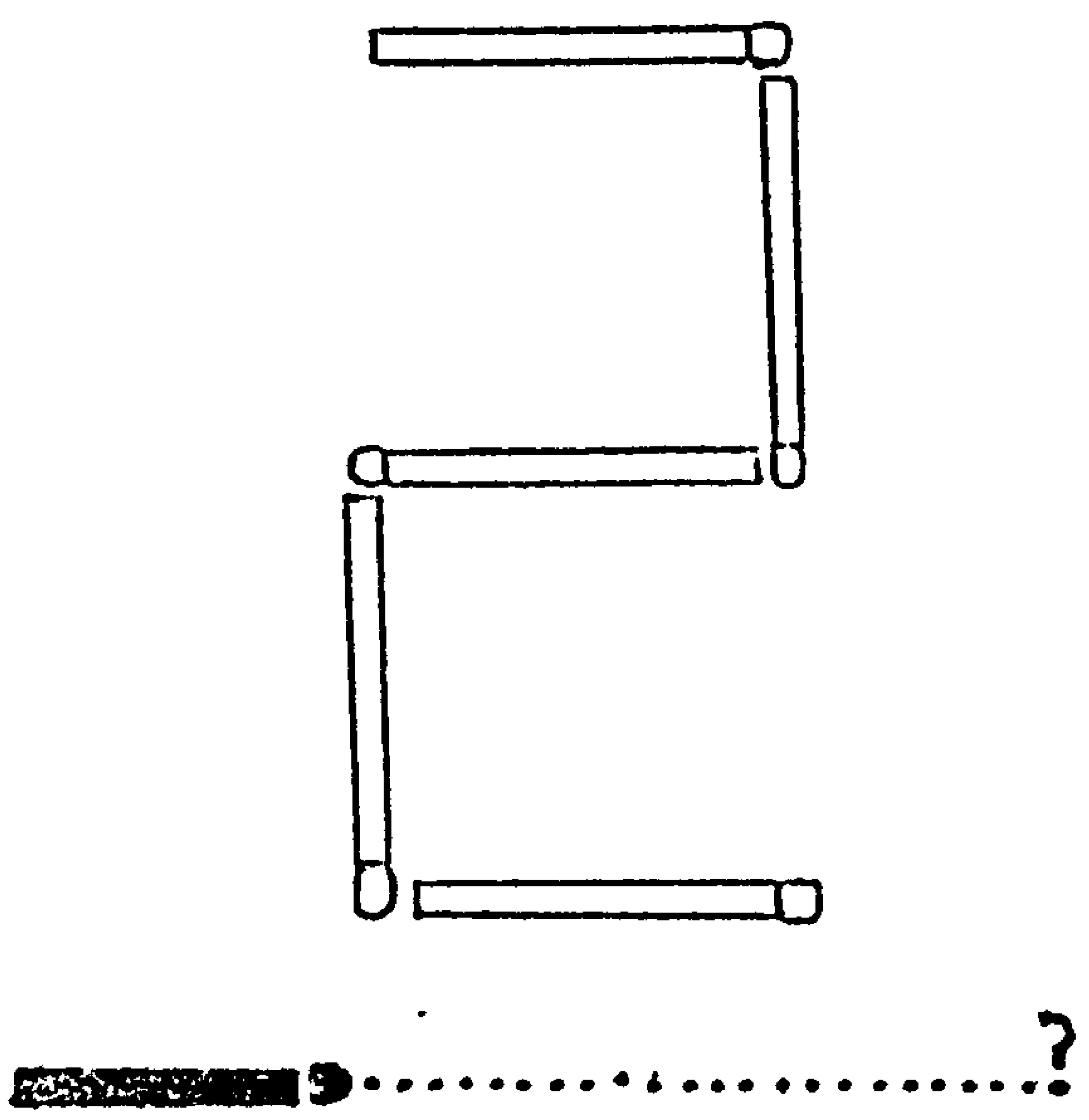
FIGURE 15: TASKS USED BY INHELDER AND SINCLAIR (1969). (AFTER INHELDER, 1972, p. 108).

A.



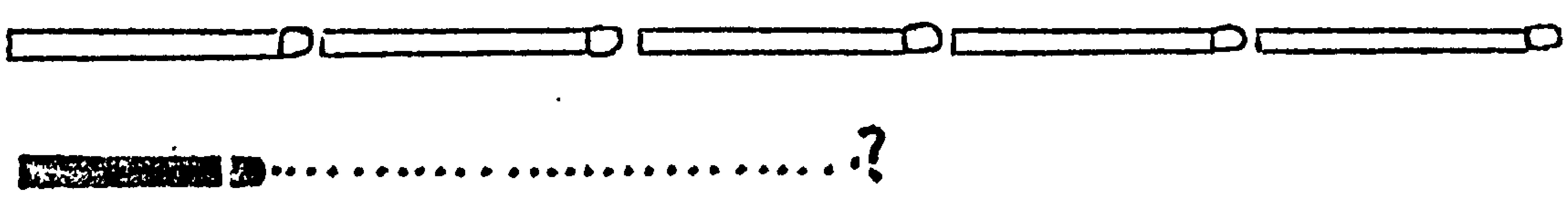
THE FIRST PROBLEM-SITUATION: CHILD MUST CONSTRUCT A STRAIGHT LINE EQUAL IN LENGTH TO THE ZIGZAG ONE.

B



THE SECOND PROBLEM-SITUATION.

C



THE THIRD PROBLEM-SITUATION.

but the subject's line is no longer directly underneath the experimenter's.

3. The experimenter constructs a straight line using one length of match, and the subject's task is to construct a straight line of the same length directly underneath it with matches of the other length.

The two lengths of matches were arranged so that the third type of problem was soluble (i.e. a simple proportional relationship existed between the two lengths). Figure 15, which is taken from Inhelder (1972), probably makes the tasks clearer.

All the subjects used by Inhelder and Sinclair were capable of conserving discontinuous quantity, but the efficacy of the experiment in training conservation of length varied from individual to individual. Four main steps in the construction process leading to attainment of the conservation were observed. Initially many subjects based their judgements on either of the main possible criteria, point-to-point length or number of matches, according to the question asked. The second step in the process seems to involve awareness of the applicability of both criteria, but inability to resolve their contradiction. This leads to a frustrating alternation between possibilities.

The third step did not always appear, but Inhelder (1972) considers it the most significant. It consists of various ingenious but inadequate attempts to integrate the two criteria, for example by breaking matches to get the same number and same point-to-point length, or by placing matches vertically instead of horizontally. This method is inadequate because it can only be used in two of the three situations. This is made clear in figure 16, which also comes from Inhelder (1972).

The explanation of the fourth step is best left to Inhelder:

'From here, the fourth step in the construction process follows for many subjects. Instead of one scheme operating a post-hoc correction on the other, we now see a reciprocal adjustment whereby the criterion of coincidence (sufficient provided the two paths are parallel) and that of numerical equality (sufficient provided all matches are of equal length) are successfully integrated into a coherent system which allows the child to solve problems of length in all generality, and no longer in special cases. Now the different schemes can be integrated, which gives a new impetus to the search for necessary and sufficient conditions for equality of length. This results in a complete understanding of the compensation involved. The children explain "You need more matches when they are smaller" and "the path goes less far but it has zig-zags".' (Inhelder, 1972, pp. 110-111. Underlined passages italicised in original).

It is most interesting to compare the result of having Q_c and Q_c available in this situation with the result of the modified roulette experiment. In both cases the child is faced with the difficult task of reconciling conflicting ways of making quantity judgements. Of course, in the modified roulette experiment size estimation and numbers of pieces are not always conflicting cues, and the children have no way of making the ingenious compromises described by Inhelder. Nevertheless, some subjects are clearly disturbed by the instances where utilising the stage 2 strategy with numbers of pieces or sizes of segments as the feature to be quantified leads to opposite choices, and the amount of switching between the two cues confirms that most subjects do not continually favour one or the other.

The ages of the subjects involved in the experiments reported by Inhelder are presumably below those of the subjects alternating between number and size cues in the modified roulette experiment (the mean age quoted by Inhelder and Sinclair, 1969, is six years), and those in the experiments of Piaget (1952) and Bryant (1971b) are younger still (roughly 5-6 for Piaget, and 3-4 for Bryant). However, this does not seriously detract from their interest in relation to the modified roulette results, as the situation used in the modified roulette experiment is more complex, so that an effect of the type Piaget calls a 'décalage' (see Flavell, 1963) would be expected.

<u>Some final theoretical considerations and speculations</u>.

The conclusion Inhelder (1972) draws from the results of the Genevan training studies is as follows:

> 'Instead of a more or less straightforward type of development, with differentiations becoming more and more refined (in the form of a treelike diagram), the interactions between different subsystems appear of the greatest importance. As the first example has shown, interactions between numerical and ordinal ways of dealing with problems of judging or constructing lengths lead to a conflict. It is this conflict which will trigger the process leading to the final resolution, through reciprocal assimilation of the two different subsystems that do not necessarily belong to the same developmental level'. (Op.cit., p. 113).

The similarity of this theoretical view to the position developed here can be clearly seen, despite the differences in terminology. Inhelder then shows how she thinks her 'process model' relates to the 'classical Piagetian structural model'. This is not an impossible task for, as Newell (1972) makes clear, the terms structure and process can only be used in relation to an implied perspective. In other words, what is structure-like and what is process-like depends on the viewpoint of the person making the distinction.

Before considering Inhelder's views on this matter, an attempt will be made to surmount some of the limitations of her position. These arise from the fact that the cases she has considered are in many ways rather special, and only shed light on a small part of cognitive development. Of course, this is not a criticism, but a natural consequence of any carefully arranged and controlled investigation. Hence some profit may be derived from fitting her detailed studies into a more general framework.

In order to broaden our understanding of children's performance in experiments it is necessary to consider not just the strategies they use but the range of possible strategies from which these are selected. Such an approach is outlined by Goodnow (1972), who stresses two themes in particular; performance as selection from a range of possible behaviours, and learning as the development of knowledge about possible, usual, and most appropriate selections for a given situation (Goodnow, 1972, p. 96).

One of the main aims of the traditional experimental method is, of course, to eliminate this selective nature of performance by controlling conditions to the point where we can hopefully discover the presence or absence of a single item in the child's repertoire. Although this will probably never be achieved in practice, it remains for many researchers as an ideal to be worked towards. Others see it as narrowly restricting and urge the analysis of data from non-experimental 'realistic' settings (e.g. Hutt and Hutt, 1970), but this merely leads to different shortcomings. The interesting thing about Goodnow's (1972) position is that it forms a kind of link between the two approaches, because the obvious thing to do if one accepts it is to study both the particular strategies children use and the total repertoire from which these are drawn.

Such an approach leads to many questions for future work following from the present research. A particularly interesting example concerns the availability to the child of different strategies in a developmental sequence. It has been shown that the same child can use different strategies in different situations, and even different strategies in similar situations of differing levels of complexity. It would be interesting to know to what extent the children are themselves conscious of this, to what extent it is only typical of transition periods, how far they can'regress' or'advance' in appropriate conditions, and so on. Above all, it is essential to discover how many ways of approaching a problem children see as viable.

Another important task stemming from the idea of performance as selection from a repertoire is to investigate the criteria on which such a selection is based. A small project of this type has already been outlined for part of the repertoire seen in the results of the modified roulette experiment, and this might serve as the foundation for a broader treatment of the problem. Other ideas concerning the social constraints of the experimental situation have also been put forward, and his interrpretation of such constraints obviously plays an important role in the subject's choice of behaviour. Goodnow makes this point with examples borrowed from Arnheim (1969),

> 'When a teacher or experimenter says "this is a test", or, even more ambiguously, "I have some games for you to play", the expectation is that the subject will respond by putting out a fair amount of effort. The subject should also know that the expected answers are "scientific" ones........"scientific" answers are expected when someone asks, as part of the Wechsler scale, "How are wood and alcohol alike?" The answer, "They both knock you out" may be regarded as witty in some settings, Arnheim points out, but it will rate you zero points in a testing situation. "Good" "scientific" answers, of course - e.g., "They are both hydrocarbons" - will rate you zero points in some other conversations'. (Goodnow, 1972, p. 98.)

An important theoretical question concerns the origin and status of the repertoire of strategies from which performance is drawn. Goodnow seems to take the view that the strategies are held in a 'made-up' form, ready to be applied whenever conditions are appropriate. This means that development is seen as the addition of strategies to the repertoire and the elaboration of the criteria for applying strategies. A similar position is adopted by Miller, Galanter, and Pribram (1960),

> 'Where do Plans come from? Probably the major source of new Plans is old Plans. We change them around a little bit each time we use them, but they are basically the same old Plans with minor variations. Sometimes we may borrow a new plan from someone else. But we do not often create a completely new Plan'. (Op.cit., p.177).

Later on, however, they modify this view somewhat and argue that only a 'metaplan', from which a large number of different Plans can be generated as they are needed must be stored in a person's memory.

This seems a more realistic statement, and fits some of the ideas developed here. One of the main themes of this thesis has been that although the ways in which children organise quantified information to solve problems may be equivalent over a variety of situations, their choices of cues to be quantified and quantification operator are more directly linked to particular requirements.

It is also possible, however, that strategies are not stored even in the form of metaplans, but that they are assembled specifically for the task in hand, as Klahr and Wallace (1970a) seem to imply. The truth is probably that both views have some validity. In the present case, for example, it seems quite plausible that the stage 2 strategy of comparing the amounts of whatever you want in each collection may be a general way of dealing with many problems. On the other hand, the stage 3 strategy of comparing the difference between the amount of what you want and the amount of whatever you don't want in each collection, may well be an approach that is only figured out when conditions are favourable. The results of the experiments can certainly be interpreted in this way, although it is very speculative.

All these different ways of viewing children's performance and abilities contain implicit references to ideas as to how these abilities develop. They seem, for the most part, to regard development as increasing both the size of the child's repertoire of possible strategies and the complexity of the rules he uses in selecting from this repertoire for a particular situation. In addition, some developments can be seen as resulting from modifications to existing strategies.

In this thesis special attention has been paid to the last possibility by formulating models in such a way that they can form a sequence of increasingly elaborate yet related possibilities. This is in accordance with the criterion of developmental tractability suggested by Klahr and Wallace (1970b). Evidence was then adduced that some of the models at least could be considered as equivalent to developmental stages in certain situations, but nothing has as yet been said about what might account for such a development.

It is widely believed that development is lawful, and that one of the major tasks of developmental psychology is to discover the rules underlying it. The status of such 'transition rules' varies from theory to theory. Werner (1957), for example, provides an abstract description of certain organisational features of a system and argues that in organismic development these features will tend toward better adaptation to the environment. Perhaps this can be made clearer by considering the following statment, 'It is an orthogenetic principle which states that whenever development occurs it proceeds from a state of relative globality and lack of differentiation to a state of increasing differentiation, articulation, and hierarchic integration'. (Werner, 1957, p.126).

This view of development is stated at a level of abstraction which makes it almost unfalsifiable. It is also not linked to any specification as to why development proceeds in this way, except for appeals to further abstract ideas, such as 'adaptation.' As an overall description to be linked with other, more precise ideas it may, however, be of value.

The equilibration model of Piaget (1954) is similar to Werner's views in many ways. Piaget argues that the adaptation of an organism to its environment can be understood in terms of two processes, assimilation and accomodation. These are difficult to specify concisely, but roughly speaking assimilation is used to describe the way in which the organism must fit environmental events into its own knowledge structure, and so modify and impoverish them, whilst accomodation describes the changes in internal structure which are made from time to time in order to further adaptation (this is slightly oversimplified, since a predominantly accomodative change, such as an imitation, is not adaptive in Piaget's sense). When a developmental change occurs, assimilation and accomodation will be in equilibrium.

This model draws attention to certain important aspects of developmental changes, and provides a frame of reference for studying development, but hardly explains it. Assimilation seems to be a property of any information-handling system, and the main explanatory 'work' is done by the concept of accomodation. The drawback is that the types of accomodation which are possible or not possible, or simply likely, cannot be derived directly from the model. Furthermore, the claim that assimilatory and accomodatory activity must be in equilibrium whenever a development occurs cannot be substantiated without some kind of measure of these processes. It must, however, be conceded that they do take place simultaneously whenever a development occurs for the simple reason that assimilation will always be going on and the accomodation is the development.

Many psychologists seem to prefer to approach the problem of development from less abstract levels. Bryant (1971a,b) has stressed the role of such factors as encoding and memory capacity rather than the Genevan concern with logical operations, and Pascual-Leone (1970) has advanced an explanation of this type supported by evidence that the child's memory capacity increases with increasing age (in terms of the number of 'chunks', not just more efficient 'chunking'). These are undoubtedly important influences on children's performances, and need to be worked into any satisfactory account of development but, as Flavell (1971) points out, it seems unlikely that they can carry the full burden of explaining development.

Another widely favoured notion seems to be that of cognitive conflict. Bruner (1966), for example, makes much of the conflict between modes of representation as a source of 'impulsion to grow,' although he admits that not all conflicts do lead to growth. Many examples of more or less successful training studies attempting to accelerate development by inculcating cognitive conflict have been reported by Smedslund (e.g. Smedslund, 1964, 1968).

One of the important things about Inhelder's (1972) paper is that she tries to link this kind of approach to the Genevan 'structural' perspective. As has been mentioned, she rejects the idea that all development consists simply in increasing differentiation of structures, and stresses instead the interactions between different subsystems. These subsystems can, however, be regarded as internalised schemes, so that their appearance and their possibilities of being integrated with other schemes may be determined by the general structure of the corresponding level of development.

This approach raises many interesting questions of a type which the older version of the equilibration model ignores concerning the ways in which different subsystems can be brought into line with each other and integrated into a more coherent system.

The two systemic principles on which Klahr and Wallace (1973) build up their developmental theory can help in this respect. These principles state that if the child is conceptualised as a developing information processing system:

(a) The developmental system constantly searches for consistent sequences which will enable it to eliminate redundant processing. (A consistent sequence is an internal representation of environmental inputs and system processes that always yield the same result).

(b) If, in a particular context, the system is unable to detect consistent sequences, it widens the basis of its search.

Tentative suggestions as to how some of the other questions that Inhelder's approach raises might be answered are given by Cellérier (1972), in a paper in which he discusses how the central concepts of Piaget's theory might be refined to a point where they would be programmable. Some of his ideas are worth quoting in full:

> 'The main idea seems to be that the new rules and concepts arise from recombination of the ones that are present. This

recombination relies heavily on the existence of a general purpose representation system that can code both actions and situations, and of a pattern recognition device that acts on these representations to produce rules and concepts - that is, higher order entities such as prescriptions and descriptions. This means the representation system must have some capacity to accomodate new types of input; it must itself be adaptive. Finally, there must exist a decomposition and recombination device that acts on these descriptions and prescriptions to generate new ones. The actual choice of which combinations should be generated would have to be based on the construction of a succession of partial reorganized representations (structured models) of the relationship between precriptions and descriptions, and of techniques for transcribing one into the other, thereby linking the structural and process descriptions. The selection of the adapted combinations would depend on an evaluation of their effects on the (external) problem environment, this evaluation being used to update the internal model and start a new recombination sequence'. (Op.cit., pp.120-121).

Such a view both reinforces and extends some of the conclusions which have been derived here. The idea that performance strategies are not stored in a ready-made form, but are assembled in various ways to fit situational requirements, is conveyed most forcefully, and the view that one of the keys to the puzzles of how this is achieved, and how development takes place, lies in the representations the child uses, is put forward. The empirical evidence produced here certainly cannot compel anyone to accept this interpretation, but it is consistent with it. The most important thing, however, is that the formulation is sufficiently detailed for the path to further tests to be clear, and this remains a task for the future.

SUMMARY.

A summary of the argument presented in this thesis will now be given. This summary is intended to show the way in which the themes are developed, and so is less formalised and less factual then the thesis abstract.

The study investigated the way children reason when faced with a type of problem concerning what adults would call 'chance' or 'probability'. The problems used involved two collections of elements of two colours. Children were asked which collection they would prefer in order to gamble for a specified outcome. Some of the research literature on this topic was reviewed, and four possible ways of solving such problems were suggested:

Method 1: Guessing, alternating choices, and other irrelevant methods.

Method 2: Comparing the amounts of the target elements in each collection, and choosing the collection with the greater amount.

Method 3: Comparing the differences between the amount of target and non-target elements in each collection, and choosing the collection with the most favourable difference.

Method 4: Comparing the proportions of target and non-target elements in each collection, and choosing the collection with the most favourable proportion.

Ways in which these possibilities might be investigated were discussed, and it was decided to use an experimental design which would allow the choices made to be used in inferring strategies, and to supplement this with an analysis of the reasons given by children to account for their choices. An experiment was then carried out using collections of red and green beads in boxes as materials. The subjects were 72 children aged 5 - 10 years and 48 children aged 11 - 14 years. The data obtained (choices made and reasons given for each choice) was analysed by interpretation according to specified criteria, and independently by the Brimer cluster analysis. This showed that,

(a) The way the children's solutions to the problems develop can be represented by three stages, corresponding to methods 1 - 3.

(b) Method 4 did not appear as a fourth stage with the groups of children studied.

A control experiment was performed to show that this result was not due to an artifact arising from the design of the experimental materials.

Consideration was then given to the role of counting as a convenient

and reliable method of quantification in producing this result. An experiment was designed in order to find out whether the same stages would show up in situations where only estimation was possible. For this experiment two different-sized circles were each marked into a red and green segment. There was a pointer on each circle which, when spun, would stop pointing to the red or green segment, and the subject's task was to choose one of the circles in order to 'get' a specified colour. 72 children aged 6 - 11 years acted as subjects, and the data was again analysed both by interpretation according to specified criteria and the Brimer cluster analysis. This time the findings were that,

(a) There was only a two-stage development (from method 1 to method 2) in the age range tested, since although method 3 was used by some of the subjects it never became the dominant strategy.

(b) Method 4 did not appear as a stage in the age range studied.

(c) The main methods of quantification referred to by the children were size estimation and recognition of fractions.

(d) The result appears retarded by comparison with the result of the experiment involving beads.

This result was explained by a discussion of the slower development of estimation than counting, and by reference to the imprecision of estimation. The fact that a fraction carries the same information as a properly calculated probability index, and its possible influence on the results, was also considered.

In order to test this explanation an experiment was devised in which both estimation and counting would be available. This was achieved by dividing the segments on the circles used in the second main experiment into pieces, so that subjects could estimate the sizes of the segments or count the numbers of pieces in the segments. It was predicted that counting would be preferred because of its greater precision, and that a result like that of the first main experiment (using beads) would be obtained. 60 children aged 6 - 10 years were used, and the analysis was carried out in the same way as the analyses of the other experiments. The result showed,

(a) A two stage development (from method 1 to method 2) in the age range tested.

(b) The second stage of this development was in certain respects quite stable at age 10, and there was little sign of further development to a third stage corresponding to method 3 or method 4.

(c) Wide individual differences in the methods of quantification used by the children in producing solutions to the different problems of the experiment.

(d) The rate at which the stage development took place was similar to that found in the second main experiment, not the first main experiment.

These results did not support the predictions made. It was suggested that the explanation of the result lay in a failure by the children to consistently prefer counting or estimation as a method of quantification. In order to test this view the criteria on which the choice of counting or estimation at different ages is based would have to be studied, and suggestions were made as to how this might be done. Finally, an attempt was made to fit the research findings into a wider context of other researches and theories.

APPENDIX A

Results of the preliminary experiment

Contents:

Sample responses.

Problem 1(a): One throw of a coin to get a head, or one throw of a die to get a six.

Correct choice for 'right' reason:

SP8: Chose the coin. 'It's only got two sides'.

SP5: Chose the coin. 'It's easier because there's only two, and lots of numbers on the dice.'

Correct choice for 'wrong' reason, or incorrect choice:

SP1: Chose the die. 'No reason'.

'Intuitively' correct choice:

SP3: Chose the coin, because she thought it would be 'easier'.

Problem 1(b): One throw of a coin to get a head, or two throws of a die to get a six.

Correct choice for 'right' reason:

SP5: Chose the coin. 'It's easier because there's only two, and lots of numbers on the dice'.

SP17: Chose the coin. 'It's got a fifty-fifty chance'.

Correct choice for 'wrong' reason, or incorrect choice:

SP1: Chose the die. 'I like the dice. I like the colour'.

SP2: Chose the die because she had lost with the coin.

Problem 4(a): Predicting the outcome of one toss of a coin.

Correct prediction for 'right' reason:

SP6: Predicted heads, although she said that both possible outcomes were equally likely.

Correct prediction for 'wrong' reason, or incorrect prediction:

SP1: Predicted tails. 'Because they're better. Because I haven't had it'.

Problem 4(b): Predicting the outcome of two tosses of a coin.

Correct prediction for 'right' reason:

SP15: Predicted a head and tail. 'The chances are two of the same won't come together because if you said two heads you would lose on the first go if a tail came up'.

'Intuitively' correct prediction:

SP11: Predicted a head and tail. 'It's most likely'. When asked why it was most likely he replied 'It just seems it'.

Full list of problems used in the experiment.

The problems used offered the following choices:

Problem 1:

Part (a) One throw of a die to get a six, or one toss of a coin to get a head.

Part (b) Two throws of a die to get a six, or one toss of a coin to get a head.

Part (c) Three throws of a die to get a six, or one toss of a coin to get a head. Further parts can be added as desired.

Problem 2:

Part (a) The subject is asked to predict the outcome of one throw of a die, and to give a reason for his prediction.

Part (b) The subject is asked to predict the total resulting from throwing two dice simultaneously, and to give a reason for his prediction.

Problem 3:

Part (a) A choice between getting a head with one toss of a coin, or a two, four, or six, with one throw of a die.

Part (b) A choice between getting two heads with two throws of a coin, or a six with one throw of a die.

Part (c) A choice between getting three heads with three throws of a coin, or a six with one throw of a die. Further parts can be added as desired.

Problem 4:

Part (a) The subject is asked to predict the outcome of one toss of a coin, and to give a reason for his prediction.

Part (b) The subject is asked to predict whether the outcome of two tosses of a coin will be two heads, two tails, or a head and tail, and to give a reason for his prediction.

Part (c) The subject is asked to predict whether the outcome of three tosses of a coin will be three heads, three tails, two tails and a head, or two heads and a tail, and to give a reason for his choice. (This part can be omitted for subjects unable to answer part (b) satisfactorily).

Problem 5: A choice between getting one, two, three, four, five, or six, with one throw of a die, or a head with one toss of a coin.

TABLE OF RESULTS TO PROBLEMS IN TERMS OF 'CORRECT' CHOICES.

'CORRECT' CHOICE IS DEFINED AS CHOICE OF THE GAMBLE WHICH IS MATHEMATICALLY MOST LIKELY TO LEAD TO SUCCESS. IN CASES WHERE EITHER GAMBLE IS EQUALLY LIKELY TO SUCCEED, AND THE SPECIAL CASE OF 1(C), WHERE MOST SUBJECTS ARE CONVINCED THAT EITHER CHOICE IS EQUALLY LIKELY TO LEAD TO SUCCESS, NO 'CORRECTNESS' SCORES HAVE BEEN MADE.

SUBJECT:	SP1	SP2	SP3	SP4	SP5	SP6	SP7	SP8	SP9	SP10	SP11	SP12	SP13	SP14	SP15	SP16	SP17	SP18
AGE:	5	7	7	9	9	9	10	10	11	11	12	12	13	13	14	17	17	21+
PROBLEM 1: (a)	X	X	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
(b)	X	✓	X	X	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
(c)																		
(d)		✓	✓	✓	✓	✓	✓	X		X	✓	✓	✓	✓	✓	✓	✓	✓
(e)								X		✓								
(f)								X										
PROBLEM 2: (a)																		
7 MARKED ✓✓ 6 OR 8 MARKED ✓ } (b)			✓	✓✓			✓✓	✓	✓		✓	✓	✓✓			✓✓		✓✓
PROBLEM 3: (a)																		
(b)	X	✓	✓	✓	✓	✓	✓	✓	X	✓	X	✓	✓	✓	✓	X	✓	✓
(c)	X	X	X	X	X	X	X	X	✓	✓	✓	X	X	✓	✓	✓	✓	✓
(d)	X		✓		X	X	✓	X		X		X	✓					✓
(e)					✓	✓		✓		✓		X						
(f)										✓		✓						
PROBLEM 4: (a)																		
(b)	✓	X	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
(c)					X	✓		✓	✓	✓	✓	X	✓	✓	✓		X	✓
PROBLEM 5.	✓	X	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓ = CORRECT CHOICE.

X = INCORRECT CHOICE.

TABLE OF RESULTS IN TERMS OF 'CORRECT' CHOICES FOR 'RIGHT' REASONS.

IN THE SPECIAL CASE OF PROBLEM 1(c), THE ARGUMENT THAT THERE ARE THREE TRIES WITH ONE CHANCE IN SIX OF SUCCEEDING, OR ONE TRY WITH ONE CHANCE IN TWO, HENCE BOTH ARE EQUALLY LIKELY ($3 \times \frac{1}{6} = \frac{1}{2}$), IS REGARDED AS 'CORRECT'.

SUBJECT:		SP1	SP2	SP3	SP4	SP5	SP6	SP7	SP8	SP9	SP10	SP11	SP12	SP13	SP14	SP15	SP16	SP17	SP18
AGE:		5	7	7	9	9	9	10	10	11	11	12	12	13	13	14	17	17	21+
PROBLEM 1:	(a)	X	X	I	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
	(b)	X	X	X	X	✓	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
	(c)	X	X	X	X	X	X	X	✓	X	X	?	X	✓	✓	✓	✓	✓	✓
	(d)		X	X	X	I	I	I	X		X	✓	I	✓	✓	✓	✓	✓	✓
	(e)								X		I								
	(f)								X										
PROBLEM 2:	(a)	X	?	?	?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	(b)	X	X	X	X	X	X	I	✓	X	X	I	X	X	X	X	✓	X	X
PROBLEM 3.	(a)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	?	X	✓	✓
	(b)	X	X	X	X	I	I	I	I	X	✓	X	I	I	I	I	I	✓	X
	(c)	X	X	X	X	X	X	X	X	X	X	I	X	X	I	I	X	X	X
	(d)	X		X		X	X	I	X		X		X	I					I
	(e)					I	I		X		I		X						
	(f)										I		I						
PROBLEM 4:	(a)	X	?	X	?	✓	✓	✓	✓	✓	✓	✓	?	✓	✓	✓	✓	✓	✓
	(b)	X	X	X	X	X	X	?	✓	I	I	I	I	X	X	✓	I	X	X
	(c)					X	X		✓	?	I	I	X	X	X	✓		X	X
PROBLEM 5:		✓	X	✓	✓	?	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓ = CORRECT CHOICE FOR RIGHT REASON.
X = CORRECT CHOICE FOR WRONG REASON, OR INCORRECT CHOICE.
I = INTUITIVELY CORRECT.
? = VALIDITY OF REASONING UNCERTAIN.

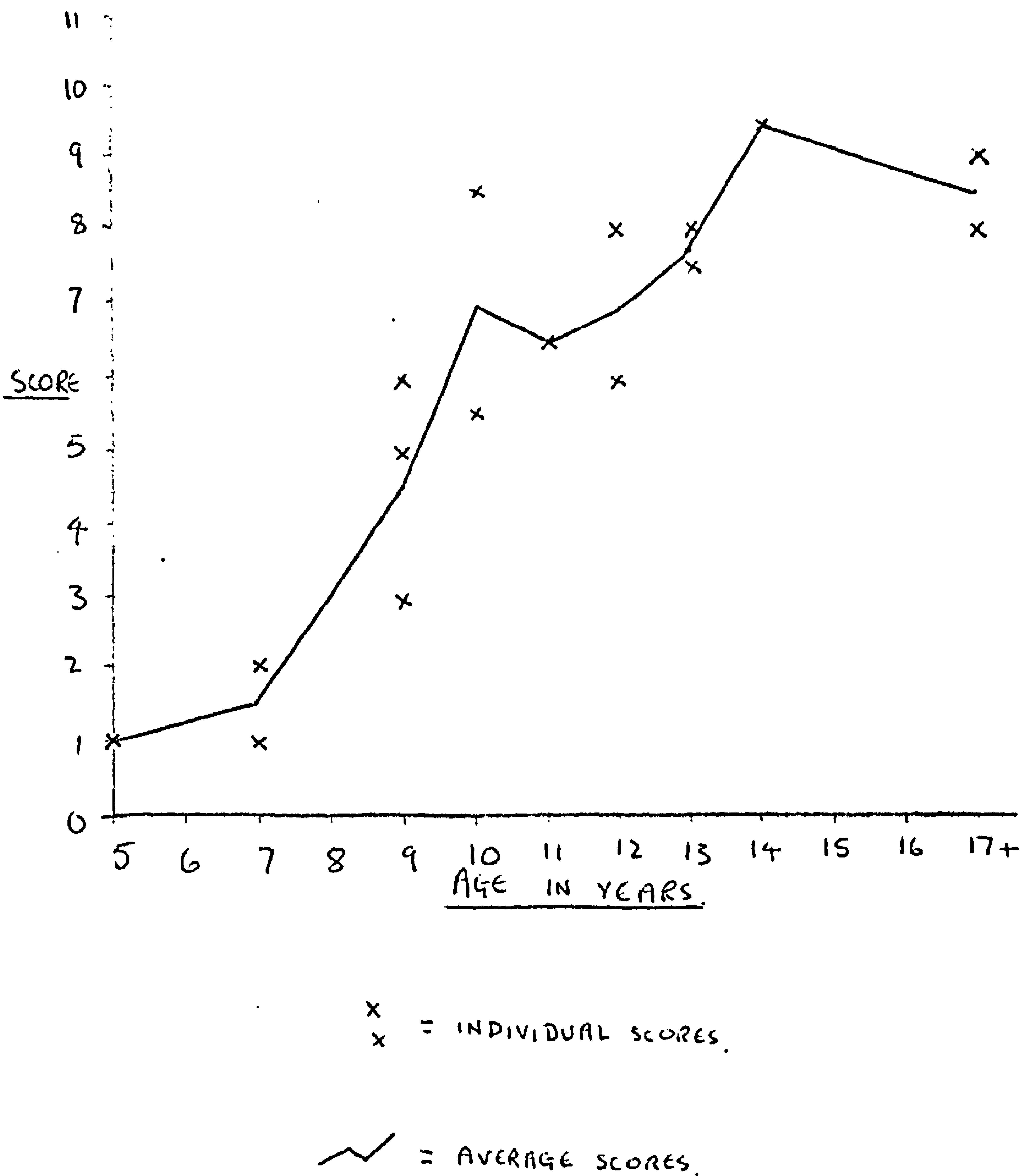
GRAPHICAL ILLUSTRATION OF THE RESULT OF THE PRELIMINARY EXPERIMEN

THIS HAS BEEN MADE BY SCORING THE ITEMS 1(a) – 1(d), 2(a), 2(b), 3(a) – 3(c), 4(a), 4(b), 5, IN THE FOLLOWING MANNER:

CORRECT REASON AND CORRECT CHOICE = 1.
UNCERTAIN REASON AND CORRECT CHOICE = $\frac{1}{2}$.
INCORRECT REASON OR INCORRECT CHOICE = 0.

THE RESULTS OF SP9 WERE DISCOUNTED ON THE GROUNDS THAT THEY WERE ABNORMALLY LOW.

THIS PRODUCES THE FOLLOWING GRAPH:



ESTIMATION OF THE CORRELATION OF ABILITY ON PROBLEMS CONCERNING PROBABILITY WITH AGE.

THIS IS OBTAINED BY MEANS OF A PRODUCT-MOMENT CORRELATION OF THE SUBJECTS' AGES AND SCORES ILLUSTRATED IN THE GRAPH. THESE ARE AS FOLLOWS:

AGE	SCORE	PRODUCT
5	1	5
7	1	7
7	2	14
9	3	27
9	5	45
9	6	54
10	5.5	55
10	8.5	85
11	6.5	71.5

AGE	SCORE	PRODUCT
12	8	96
12	6	72
13	7.5	97.5
13	8	104
14	9.5	133
17	8.5	144.5
17	9	153
21+ (COMPUTED AS 21)	8	168

CORRELATION COEFFICIENT $= r = \dfrac{\left(\Sigma XY - \dfrac{\Sigma X \Sigma Y}{n}\right)}{\sqrt{\left(\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}\right)\left(\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{n}\right)}}$

$\Sigma X = 196$, $\Sigma Y = 103$, $\Sigma XY = 1331.5$, $\Sigma X^2 = 2528$, $\Sigma Y^2 = 748.5$, $n = 17$.

$$\therefore r = \frac{\left(1331.5 - \frac{196 \times 103}{17}\right)}{\sqrt{\left[2528 - \left(\frac{38416}{17}\right)\right]\left[748.5 - \left(\frac{10609}{17}\right)\right]}} = \frac{1331.5 - 1187.5}{\sqrt{268.3 \times 124.5}}$$

$\therefore r = \underline{0.79}$ = CORRELATION OF SCORE WITH AGE

HENCE, THE Z SCORE OF R IS APPROXIMATELY 1.07.
HENCE THE RESULT IS SIGNIFICANT AT THE 0.001 LEVEL.

APPENDIX B

Interpretative analysis of results of the first beads experiment.

Contents:

TABLE OF INTERPRETATIVE CATEGORISATIONS OF BEADS EXPERIMENT RESPONSES BY SUBJECTS FROM THE PRIMARY (SB1-72) AND SECONDARY (SB73-120) SCHOOL GROUPS.

SUBJECT	AGE (YRS)	ORDER OF PRESENTATION OF PROBLEMS	PROBLEM B1		PROBLEM B2		PROBLEM B3		PROBLEM B4		PROBLEM B5		PROBLEM B6		PROBLEM B7	
			GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SB1	5	1423	1	1	1	1	1	1	1	1						
SB2	5	2413	1	1	1	1	1	1	1	1						
SB3	5	1423	1	1	1	1	1	1	1	1						
SB4	5	3241	1	1	1	1	1	1	1	1						
SB5	5	2143	1	1	1	1	1	1	1	1						
SB6	5	3124	1	1	1	1	1	1	1	1						
SB7	5	3214	1	1	1	1	1	1	1	1						
SB8	5	1432	1	1	1	1	1	1	1	1						
SB9	5	4123	1	1	1	1	1	1	1	1						
SB10	5	3241	1	1	1	1	1	1	1	1						
SB11	5	42135	1	1	1	1	2	2	2	2	2	2				
SB12	5	13245	2	2	2	2(a)	2	2	2	2	2	2				
SB13	6	2341	1	1	1	1	1	1	1	1						
SB14	6	32415	1	1	2	1	2	1	1	1	1	1				
SB15	6	2413	2	2	1	1	1	1	1	1						
SB16	6	3421	1	1	1	1	2	1	2	1						
SB17	6	41325	2	2	1?	1?	1	1?	1	2	1?	1				
SB18	6	14235	2	2	1	1	2	2	2	?	2	2				
SB19	6	14235	2	2	2	1	2	2	2	2	2	2				
SB20	6	4231	2	2	2	2(a)	2	?	2	?						
SB21	6	21435	2	2	2	2(a)	2	2	2	2	2	2				
SB22	6	14325	2	2	2	2(a)	2	2	2	2	2	2				
SB23	6	32415	2	2	2	2(b)	2	2	2	2	2	2				
SB24	6	23415	2	2	2	3?	2	2	2	2	2	2				
SB25	7	2431	1	1	1	1	1	1	1	1						
SB26	7	1423	1	1	1	1	1	1	1	1						
SB27	7	1423	1	1	2	?	?	?	2	2						
SB28	7	43215	2	2	2	2(a)	2	2	2	2	2	2				
SB29	7	24315	2	2	2	2(a)	2	2	2	2	2	2				
SB30	7	32415	2	2	2	2(a)	2	2	2	2	2	2				
SB31	7	14325	2	2	2	2(a)	2	2	2	2	2	2				
SB32	7	42135	2	2	2	2(b)	2	2	2	2	2	2				
SB33	7	43125	2	2	2	2(b)	2	2	2	2	2	2				
SB34	7	24315	3	3	2	3	2	?	?	?	3	?				
SB35	7	3142567	3	3	3	3	2	?	3	?	2	2	2	2	2	2
SB36	7	4312567	3	3	3	3	2	2	2	2	3	2	2?	2	3	2

(CONTINUED OVERLEAF)

BEADS EXPERIMENT RESULTS (CONT'D).

SUBJECT	AGE (YRS)	ORDER OF PRESENTATION OF PROBLEMS	PROBLEM B1		PROBLEM B2		PROBLEM B3		PROBLEM B4		PROBLEM B5		PROBLEM B6		PROBLEM B7	
			GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SB37	8	2314	1	1	1	1	1	1	2	1						
SB38	8	42135	2	2	2	2(a)	2	2	2	2	2	2				
SB39	8	4231	2	2	2	2(a)	2	2	2	2						
SB40	8	13425	2	2	2	2(a)	2	2	2	2	2	2				
SB41	8	12435	2	2	2	2(a)	2	2	2	2	2	2				
SB42	8	23145	2	2	2	2(a)	2	2	2	2	2	2				
SB43	8	32145	2	2	2	2(a)	2	2	2	2	2	2				
SB44	8	142335	2	2	2	2(b)	3	3	2	2	3	2				
SB48	8	423167	3	3	2	2(b)	3	3	2	3			3	3	3	3
SB45	8	132467	2	2	3	3	3	3	3	3			3	2	3	2
SB46	8	13425	3	3	3	3	3	3	3	2	3	3				
SB47	8	241367	3	3	3	3	3	3	3	3			2	2	3	3
SB49	9	2314	1	1	1	1	1	1	1	1						
SB51	9	2431567	1	1	2	2(a)	2	?	2	?	2	2	2	2	2	2
SB50	9	3142	2	2	2	?	2	?	2	1						
SB52	9	32415	2	2	2	2(a)	2	2	2	2	2	2				
SB53	9	31425	2	2	2	2(b)	2	2	2	2	2	2				
SB55	9	324167	3	3	?	?	2	?	3	3			2	?	2	2
SB54	9	2413567	3	2	2	2(b)	3	3	2	?	3	3	2	2	2	?
SB56	9	1324567	3	3	3	3	2	2	3	2	3	2	?	2	2	2
SB57	9	143267	3	3	3	3	2	2	3	3			2	2	2	2
SB58	9	342167	3	3	3	3	2	2	3	3			?	2	3	3
SB59	9	231467	3	3	3	3	2	2	3	3			2	3	3	3
SB60	9	1243567	3	3	3	3	3	3	3	2	3	3	3	3	3	3
SB61	10	324 5	2	2	2	2(a)	2	2	2	2	2	2				
SB62	10	32145	2	2	2	2(a)	2	2	2	2	2	2				
SB63	10	23415	2	2	2	2(b)	2	2	2	2	2	2				
SB64	10	32145	2	2	?	?	2	?	2	2	3	3				
SB71	10	[illegible]	2	2	3	3	3	3	2	2	3	3	3	3	3	3
SB65	10	231457	3	3	3	3	2	2	3	2	3	2	?	2	3	3
SB66	10	[illegible]	3	3	3	3	?	3	3	2	3	3	2	2	3	3
SB67	10	[illegible]	3	3	3	3	3	3	3	3			?	?	?	?
SB68	10	[illegible]	3	3	3	3	3	3	3	3			?	2	3	3
SB69	10	124367	3	3	3	3	3	3	3	3			3	?	3	3
SB70	10	412367	3	3	3	3	3	3	3	3			3	2	3	3
SB72	10	[illegible]	3	3	3	3	3	3	3	3			3	3	3	3
SB73	11	31245	2	2	2	2(a)	2	2	2	2	2	2				
SB74	11	14325	2	2	2	2(a)	2	2	2	2	2	2				
SB75	11	12435	2	2	2	2(a)	2	2	2	2	2	2				
SB76	11	42135	2	2	2	2(b)	2	2	2	2	2	2				
SB77	11	31425	2	2	2	3	2	2	2	2	2	2				
SB78	11	2143567	3	3	3	3	2	2	3	2	2	2	?	2	3	3

BEADS EXPERIMENT RESULTS (CONCLUDED).

SUBJECT	AGE (YRS)	ORDER OF PRESENTATION OF PROBLEMS	PROBLEM B1		PROBLEM B2		PROBLEM B3		PROBLEM B4		PROBLEM B5		PROBLEM B6		PROBLEM B7	
			GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SB79	11	234167	3	3	3	3	2	2	3	3			?	2	3	3
SB80	11	314267	3	3	3	3	3	3	3	2			2	2	3	2
SB81	11	[illegible]	3	3	3	3	?	?	3	3			3	3	3	3
SB82	11	[illegible]	3	3	3	3	3	3	3	3			3	3	3	?
SB83	11	[illegible]	3	3	3	3	3	3	3	3			3	3	3	3
SB84	11	324167	3	3	3	3	3	3	3	3			4	2	3	3
SB85	12	[illegible]	2	2	2	2 ?	2	2	2	2	2	2				
SB86	12	4321	2	2	2	2(a)	2	2	2	2						
SB87	12	24315	2	2	2	2(a)	2	2	2	2	2	2				
SB88	12	43215	2	2	2	2(a)	2	2	2	2	2	2				
SB89	12	2431	2	2	2	2(b)	2	?	2	3						
SB95	12	1432567	2	2	3	3	2	2	?	3	3	3	3	2	3	3
SB94	12	243167	2	2	2	2(b)	3	3	3	3			?	2	2	?
SB90	12	143267	3	3	3	2(b)	2	2	2	3			?	2	2	2
SB91	12	312467	3	3	3	3	2	2	3	2			3	3	2	2
SB92	12	3214567	3	3	3	3	2	2	3	2	3	2	2	2	3	2
SB93	12	312467	3	3	3	3	?	2	3	3			?	?	?	3
SB96	12	421367	3	3	3	3	3	3	3	3			3	3	3	4
SB97	13	1423	?	?	?	?	?	?	?	?						
SB98	13	12345	2	2	2	?	2	2	2	2	2	2				
SB99	13	12435	2	2	2	2(a)	2	2	2	2	2	2				
SB100	13	43215	2	2	2	2(b)	2	2	2	2	2	2				
SB101	13	142367	3	3	2	3	?	?	2	3			2	2	2	3
SB106	13	321467	3	3	2	3	2	2	3	3			3	2	3	3
SB102	13	314267	3	3	3	3	?	?	3	3			?	?	3	3
SB103	13	321467	3	3	3	3	?	?	3	3			?	2	3	3
SB104	13	234167	3	3	3	3	?	?	3	3			?	2	3	3
SB105	13	[illegible]	3	3	3	3	2	2	3	3			2	2	2	2
SB107	13	321467	3	3	3	3	3	3	3	3			?	2	3	3
SB108	13	134267	4	4	4	3	3	3	4	4			4	4	3	3
SB109	14	3142	2	2	2	2(a)	2	2	2	2						
SB110	14	[illegible]	2	2	2	2(a)	2	2	2	2	2	2				
SB111	14	1234	2	2	2	3	2	2	2	3						
SB112	14	1432567	3	3	3	3	2	2	3	2	3	2	2	2	2	2
SB113	14	341267	3	3	3	3	?	?	3	3			?	?	3	3
SB114	14	[illegible]	3	3	3	3	3	3	3	2			3	3	?	?
SB115	14	[illegible]	3	3	3	3	3	3	?	3	4	?	?	2	3	?
SB116	14	[illegible]	3	3	3	3	3	3	3	3			2	2	3	2
SB117	14	[illegible]	3	3	3	3	3	3	3	3			?	3	3	?
SB118	14	321467	3	3	3	3	3	3	3	3			?	2	3	3
SB119	14	341267	3	3	3	3	3	3	3	3			3	3	3	3
SB120	14	432167	3	3	3	3	3	3	3	3			3	3	4	4

TABLE OF THE FREQUENCY OF THE DIFFERENT TYPES OF RESPONSE AT DIFFERENT AGES IN THE PRIMARY SCHOOL SUBJECTS:

AGE GROUP	NUMBER OF TYPE 1 RESPONSES	NUMBER OF TYPE 2 RESPONSES	NUMBER OF TYPE 3 RESPONSES	NUMBER OF TYPE 4 RESPONSES	NUMBER OF 2(a) RESPONSES	NUMBER OF 2(b) RESPONSES	UNCLASSIFIABLE RESPONSES
5	84	12	0	0	1	0	0
6	34	58	1	0	3	1	3
7	18	58	12	0	4	2	8
8	7	61	28	0	6	2	0
9	11	40	37	0	2	2	8
10	0	37	55	0	2	1	4

TABLE OF THE FREQUENCY OF THE DIFFERENT TYPES OF RESPONSE AT DIFFERENT AGES IN THE SECONDARY SCHOOL SUBJECTS:

AGE GROUP	NUMBER OF TYPE 1 RESPONSES	NUMBER OF TYPE 2 RESPONSES	NUMBER OF TYPE 3 RESPONSES	NUMBER OF TYPE 4 RESPONSES	NUMBER OF 2(a) RESPONSES	NUMBER OF 2(b) RESPONSES	UNCLASSIFIABLE RESPONSES
11	0	45	49	0	3	1	2
12	0	57	36	0	3	3	3
13	0	20	44	5	1	1	17
14	0	26	67	0	2	0	3

APPENDIX C

Cluster analysis of results of the first beads experiment.

Contents: Page:

TABLE OF INITIAL CATEGORISATIONS OF RESPONSES TO THE BEADS EXPERIMENT ITEMS BY SUBJECTS FROM THE PRIMARY (SB1-72) AND SECONDARY (SB73-120) SCHOOL GROUPS.

SUBJECT	AGE (YRS).	PROBLEM B1		PROBLEM B2		PROBLEM B3		PROBLEM B4	
		GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SB1	5	FU	FU	PU	PU	FUL	FUM	FU	FU
SB2	5	PSR	PSR	FSR	FSR	FUM	FSRL	FQS	FSR
SB3	5	FU	FU	PU	PU	FUM	FUM	PU	PU
SB4	5	PU	PU	PU	PU	FUM	FUL	FU	FU
SB5	5	FSR	FSR	PSR	PSR	FSRL	FSRM	PSR	PU
SB6	5	PU	PPR	PRF	PRF	FUM	FUL	FPR	FPR
SB7	5	PU	PU	PU	PU	FUL	FUM	PU	PU
SB8	5	PSR	PSR	PSR	PSR	FSRL	FSRM	PSR	FSR
SB9	5	PRF	PU	FRF	PPR	FPRL	FPRM	FRF	FRF
SB10	5	FQS	FQS	FSR	FQS	FSRM	FSRL	FSR	FQS
SB11	5	FU	FU	FU	FU	FQTM	FQTM	FU	FPR
SB12	5	PQTA	PQTA	PQTA	PQTA	FQTA	FQTA	PQTA	FQTA
SB13	6	FQA	FU	FSR	FQS	FSRL	FQAL	PQS	FQS
SB14	6	PQS	PQS	PQTM	PQS	FQTM	FQSL	FQS	PQS
SB15	6	PQTM	PQTM	FQA	FQA	FQAM	FQAL	FQA	FQANT
SB16	6	PU	FU	PRF	PRF	FUL	FQTM	PQTM	PQRF
SB17	6	PQA	PQTM	PQA	PQTA	FQAM	FQAL	FQA	FQTM
SB18	6	PQTM	PQTM	FQTL	FU	FQTM	FQTM	PQTM	PU
SB19	6	PQA	PQA	PQA	FQTA	FQAM	FQAM	PQA	FQTA
SB20	6	PQTM	PQTM	PQTM	FQTM	FQTM	FQTM	PQTM	PQTL
SB21	6	PQA	PQA	PQA	PQA	FQTM	FQTM	PQA	FQTM
SB22	6	PQTM	PQTM	PQTM	FQTM	FQTM	FQTM	PQTM	FQTM
SB23	6	PQTM	PQTM	PQTM	PQNTL	FQTM	FQTM	PQTM	FQTM
SB24	6	PQTM	PQTM	PQTM	PQRM	FQTM	FQTM	PQTM	FQTM
SB25	7	PSR	PSR	FSR	FSR	FSRL	FSRM	PSR	PSR
SB26	7	PQS	PSR	PQS	PQS	FQSL	FQSM	FQS	FQS
SB27	7	FSR	PSR	PSR	FSR	FLL	FLL	PQTM	FQTM
SB28	7	PQTA	PQTA	PQTM	FQTA	FQTA	FQTA	PQTA	FQTA
SB29	7	PQTA	PQTA	PQTA	PQTA	FQTA	FQTA	PQTA	FQTA
SB30	7	PQTM	PQTM	PQTM	PQTA	FQTM	FQTM	PQTM	FQTM
SB31	7	PQTM	PQTM	PQTM	PQTA	FQTM	FQTM	PQTM	FQTM
SB32	7	PQTM	PQTM	PQTM	PQNTL	FQTM	FQTM	PQTM	FQTM
SB33	7	PQTM	PQTM	PQTM	PQNTL	FQTM	FQTM	PQTM	FQTM
SB34	7	PQRM	PQRM	PQTM	PQRA	FMM	FLL	FQA	FQRA
SB35	7	PQRM	PQRM	PQRM	PQRM	FMM	FLL	PQRM	PLL
SB36	7	PQRM	PQRM	PQRM	PQRM	FQTM	FQTM	PQTM	FQTM

(CONTINUED OVERLEAF)

INITIAL CATEGORISATIONS FOR CLUSTER ANALYSIS (CONT'D).

SUBJECT	AGE (YEARS)	PROBLEM B1		PROBLEM B2		PROBLEM B3		PROBLEM B4	
		GREN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SB37	8	PPR	PPR	PSR	FPR	FPRM	FSRM	PQTM	PU
SB38	8	PQTM	PQTM	PQTM	PQTM	FQTM	FQTM	PQTM	FQTM
SB39	8	PQTM	PQTM	PQTM	PQTA	FQTM	FQTM	PQTM	FQTM
SB40	8	PQTM	PQTM	PQTM	PQTM	FQTM	FQTM	PQTM	FQTM
SB41	8	PQTM	PQTM	PQTM	FQTM	FQTS	FQTM	PQTS	FQTS
SB42	8	PQTM	PQTM	PQTM	FQTM	FQTM	FQTM	PQTM	FQTM
SB43	8	PQTM	PQTM	PQTA	FQTA	FQTA	FQTA	PQTM	FQTM
SB44	8	PQTM	PQTM	PQTM	PQNTL	PQRM	PQRM	PQTM	FQTM
SB45	8	PQRA	PQRA	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM
SB46	8	PQROA	PQROA	PQRA	PQRA	FQTA	FQTA	PQRA	FQTA
SB47	8	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRA	PQRM
SB48	8	PQRA	PQRA	PQTM	PQNTL	PQRM	PQRM	PQTM	PQRM
SB49	9	FQS	PQS	PQS	FQS	FPRM	FPRM	FSR	FSR
SB50	9	PQTM	PQTM	PQTM	PU	FQTL	FQTM	PQTM	PU
SB51	9	FU	FU	PQTM	PQTM	FPRL	FQTM	PQTM	PQTL
SB52	9	PQTM	PQTM	PQTM	PQTM	FQTM	FQTM	PQTM	FQTM
SB53	9	PQTM	PQTM	PQTM	PQNTL	FQTM	FUM	PQTM	FQTM
SB54	9	PQNTL	PQRM	PQTM	PQNTL	PQRM	PQRM	PQTM	PQRA
SB55	9	PQRM	PQRM	FQTL	FQRM	FUM	FQTL	PQRM	PQRM
SB56	9	PQRM	PQRM	PQRM	PQRM	FQTA	FQTA	PQRM	FQTM
SB57	9	PQRM	PQRM	PQRM	PQRM	FQTM	FQTM	PQRM	PQRM
SB58	9	PQRM	PQRM	PQRM	PQRM	FQTM	FQTM	PQTM	PQRM
SB59	9	PQRM	PQRM	PQRM	PQRM	FQTM	FQTM	PQRM	PQRM
SB60	9	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	FQTM
SB61	10	PQTM	PQTM	PQTM	PQTM	FQTM	FQTM	PQTM	FQTM
SB62	10	PQTM	PQTM	PQTM	FQTM	FQTM	FQTM	PQTM	FQTM
SB63	10	PQTR	PQTR	PQTS	PQNTA	FQTR	FQTR	PQTR	FQTR
SB64	10	PQNTL	PQNTL	FQTL	PQNTL	FQTM	FQTL	PQTM	FQTM
SB65	10	PQRM	PQRM	PQRM	PQRM	FQTM	FQTM	PQRM	FQTM
SB66	10	PQRA	PQRA	PQRA	PQRA	FQTL	PU	PQRS	FQTM
SB67	10	PQRS	PQRS	PQRM	PQRM	PQRM	PQRM	PQRS	PQRM
SB68	10	PQRS	PQRS	PQRS	PQRM	PQRM	PQRM	PQRM	PQRM
SB69	10	PQA	PQANT	PQRA	PQRA	PQRM	PQRM	PQRS	PQRM
SB70	10	PQRM	PQRM	PQRS	PQRS	PQRM	PQRM	PQRM	PQRM
SB71	10	PQTM	PQTM	PQRM	PQNTL	PQRM	PQRM	PQTM	FQTM
SB72	10	PQA	PQA	PQRM	PQRM	PQRM	PQRM	PQRS	PQRM
SB73	11	PQTM	PQTM	PQTM	FQTA	FQTA	FQTA	PQTM	FQTM
SB74	11	PQTM	PQTM	PQTM	PQTA	FQTM	FQTM	PQTM	FQTM
SB75	11	PQTM	PQTM	PQTM	FQTM	FQTM	FQTM	PQTM	FQTM
SB76	11	PQTM	PQTM	PQTM	PQNTL	FQTM	FQTM	PQTM	FQTM
SB77	11	PQTM	PQTM	PQTM	PQRM	FQTM	FQTM	PQTM	FQTM
SB78	11	PQRM	PQRS	PQRM	PQRM	FQTM	FQTM	PQRM	FQTM

INITIAL CATEGORISATIONS FOR CLUSTER ANALYSIS (CONCLUDED).

SUBJECT	AGE (YRS).	PROBLEM B1		PROBLEM B2		PROBLEM B3		PROBLEM B4	
		GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SB79	11	PQRM	FQRM	PQRM	PQRM	FQTM	FQTM	PQRM	PQRM
SB80	11	PQRM	PQRM	PQRS	PQRS	PQRM	PQRM	PQRS	FQTM
SB81	11	PQRM	PQRM	PQRM	PQRS	FUL	FMM	PQRS	FQRM
SB82	11	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM
SB83	11	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM
SB84	11	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM
SB85	12	PQTM	PQTM	PQTM	FQNTM	FQTM	FQTM	PQTM	FQTM
SB86	12	PQTM	PQTM	PQTM	FQTM	FQTM	FQTM	PQTM	FQTM
SB87	12	PQTM	PQTS	PQTM	FQTM	FQTS	FQTS	PQTM	FQTS
SB88	12	PQTM	PQTM	PQTS	FQTM	FQTM	FQTM	PQTM	FQTS
SB89	12	PQTM	PQTM	PQTM	PQNTL	FQTM	FRFL	PQTM	PQRS
SB90	12	PQRA	PQRA	PQRM	PQNTL	FQTS	FQTS	PQTM	PQRS
SB91	12	PQRM	PQRM	PQRM	PQRM	FQTM	FQTM	PQRM	FQTM
SB92	12	PQRM	PQRM	PQRM	PQRM	FQTM	FQTM	PQRM	FQTA
SB93	12	PQRM	PQRM	PQRM	PQRM	FLL	FMM	PQRS	PQRM
SB94	12	PQRS	PQRM	PQTM	PQRA	PQRM	PQRM	PQRS	PQRS
SB95	12	PQTM	PQTM	PQRM	PQRM	FQTM	FQTM	FQNTL	PQRM
SB96	12	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM
SB97	13	U	U	U	U	U	U	U	U
SB98	13	PQTM	PQTM	PQTM	PPR	FQTM	FQTM	PQTM	FQTM
SB99	13	PQTM	PQTM	PQTM	FQTM	FMM	FMM	PQTM	FQTM
SB100	13	PQTM	PQTM	PQTM	PQNTL	FMM	FQTM	PQTM	FQTS
SB101	13	PQRM	PQRM	PQTM	PQRS	FQTL	FQTL	PQTM	PQRM
SB102	13	PQRM	PQRM	PQRM	PQRS	FQTL	FQTL	PQRM	PQRM
SB103	13	PQRM	PQRM	PQRM	PQRM	FQTL	FQTL	PQRS	PQRM
SB104	13	PQRM	PQRM	PQRM	PQRM	FLL	FLL	PQRM	PQRM
SB105	13	PQRM	PQRM	PQRS	PQRM	FQTA	FQTM	PQRM	PQRM
SB106	13	PQRM	PQRM	PQTM	PQRA	FQTM	FQTM	PQRM	PQRM
SB107	13	PQRS	PQRS	PQRA	PQRA	PQRM	PQRM	PQRS	PQRS
SB108	13	PQRR	PQRR	PQRM	PQRR	PQRM	PQRM	PQRR	PQRM
SB109	14	PQTM	PQTM	PQTA	FQTA	FQTM	FQTM	PQTA	FQTM
SB110	14	PQTA	PQTA	PQTA	FQTA	FQTA	FQTA	PQTM	FQTA
SB111	14	PQTM	PQTM	PQTM	PQRM	FQTM	FQTM	PQTM	PQRM
SB112	14	PQRM	PQRM	PQRM	PQRM	FQTM	FQTM	PQRM	FQTM
SB113	14	PQRM	PQRM	PQRM	PQRM	FLL	FLL	PQRM	PQNTL
SB114	14	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	FQTM
SB115	14	PQRM	PQRR	PQRM	PQRM	PQRM	PQRM	FQNTL	PQRM
SB116	14	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM
SB117	14	PQRM	PQRM	PQRS	PQRA	PQRA	PQRA	PQRM	PQRS
SB118	14	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM
SB119	14	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM
SB120	14	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM	PQRM

CATEGORY NUMBERING SCHEME FOR ANALYSIS OF RESULTS FROM THE PRIMARY SCHOOL GROUP ONLY (FIRST BEADS EXPERIMENT)

PROBLEM B1				PROBLEM B2				PROBLEM B3				PROBLEM B4			
GREEN		RED		GREEN		RED		GREEN		RED		GREEN		RED	
CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER
FU	1	FU	18	FU	34	FU	50	FUL	71	FUL	87	FU	103	FU	120
PU	2	PU	19	PU	35	PU	51	FUM	72	FUM	88	PU	104	PU	121
FSR	3	FSR	20	FSR	36	FSR	52	FSRL	73	FSRL	89	FSR	105	FSR	122
PSR	4	PSR	21	PSR	37	PSR	53	FSRM	74	FSRM	90	PSR	106	PSR	123
FQS	5	FQS	22	PQS	38	FQS	54	FQSL	75	FQSL	91	FQS	107	FQS	124
PQS	6	PQS	23	FRF	39	PQS	55	FPRL	76	FQSM	92	PQS	108	PQS	125
PPR	7	PPR	24	PRF	40	FPR	56	FPRM	77	FPRM	93	FPR	109	FPR	126
PRF	8	PQA	25	FQA	41	PPR	57	FQAM	78	PU	94	FRF	110	FRF	127
FQA	9	PQANT	26	PQA	42	PRF	58	FLL	79	FQAL	95	FQA	111	PQRF	128
PQA	10	PQTA	27	PQTA	43	FQA	59	FMM	80	FQAM	96	PQA	112	PLL	129
PQTA	11	PQTM	28	PQTM	44	FQTA	60	FQTL	81	FLL	97	PQTA	113	FQANT	130
PQTM	12	PQTR	29	PQTS	45	FQTM	61	FQTA	82	FQTL	98	PQTM	114	FQTA	131
PQTR	13	PQNTL	30	FQTL	46	PQA	62	FQTM	83	FQTA	99	PQTS	115	FQTM	132
PQNTL	14	PQRA	31	PQRA	47	PQTA	63	FQTS	84	FQTM	100	PQTR	116	FQTS	133
PQRA	15	PQROA		PQRM	48	PQTM	64	FQTR	85	FQTR	101	PQRA	117	FQTR	134
PQROA		PQRL	32	PQRS	49	PQNTA	65	PQRM	86	PQRM	102	PQRM	118	PQTL	135
PQRM	16	PQRM				PQNTL	66					PQRS	119	FQRA	136
PQRS	17	PQRS	33			FQRM	67							PQRA	137
						PQRA	68							PQRM	138
						PQRM	69								
						PQRS	70								

DATA FOR CLUSTER ANALYSIS OF RESULTS FROM THE PRIMARY SCHOOL GROUP ONLY.

SUBJECT	CATEGORIES							
SB1	1	18	35	51	71	88	103	120
SB2	4	21	36	52	72	89	107	122
SB3	1	18	35	51	72	88	104	121
SB4	2	19	35	51	72	87	103	120
SB5	3	20	37	53	73	90	106	121
SB6	2	24	40	58	72	87	109	126
SB7	2	19	35	51	71	88	104	121
SB8	4	21	37	53	73	90	106	122
SB9	8	19	39	57	76	93	110	127
SB10	5	22	36	54	74	89	105	124
SB11	1	18	34	50	83	100	103	126
SB12	11	27	43	63	82	99	113	131
SB13	9	18	36	54	73	95	108	124
SB14	6	23	44	55	83	91	107	125
SB15	12	28	41	59	78	95	111	130
SB16	2	18	40	58	71	100	114	128
SB17	10	28	42	63	78	95	111	132
SB18	12	28	46	50	83	100	114	121
SB19	10	25	42	60	78	95	112	131
SB20	12	28	44	61	83	100	114	135
SB21	10	25	42	62	83	100	112	132
SB22	12	28	44	61	83	100	114	132
SB23	12	28	44	66	83	100	114	132
SB24	12	28	44	69	83	100	114	132
SB25	4	21	36	52	73	90	106	123
SB26	6	21	38	55	75	92	107	124
SB27	3	21	37	52	79	97	114	132
SB28	11	27	44	60	82	99	113	131
SB29	11	27	43	63	82	99	113	131
SB30	12	28	44	63	83	100	114	132
SB31	12	28	44	63	83	100	114	132
SB32	12	28	44	66	83	100	114	132
SB33	12	28	44	66	83	100	114	132
SB34	16	32	44	68	80	97	111	136
SB35	16	32	48	69	80	97	118	129
SB36	16	32	48	69	83	100	114	132

(CONTINUED OVERLEAF).

DATA FOR PRIMARY SCHOOL GROUP CLUSTER ANALYSIS (CONT'D).

SUBJECT	CATEGORIES.							
SB37	7	24	37	56	77	90	114	121
SB38	12	28	44	64	83	100	114	132
SB39	12	28	44	63	83	100	114	132
SB40	12	28	44	64	83	100	114	132
SB41	12	28	44	61	84	100	115	133
SB42	12	28	44	61	83	100	114	132
SB43	12	28	43	60	82	99	114	132
SB44	12	28	44	66	86	102	114	132
SB45	15	31	48	69	86	102	118	138
SB46	15	31	47	68	82	99	117	131
SB47	16	32	48	69	86	102	117	138
SB48	15	31	48	66	86	102	114	138
SB49	5	23	38	54	77	93	105	122
SB50	12	28	44	51	81	100	114	121
SB51	1	18	44	64	76	100	114	135
SB52	12	28	44	64	83	100	114	132
SB53	12	28	44	66	83	88	114	132
SB54	14	32	44	66	86	102	114	137
SB55	16	32	46	67	72	98	118	138
SB56	16	32	48	69	82	99	118	132
SB57	16	32	48	69	83	100	118	138
SB58	16	32	48	69	83	100	114	138
SB59	16	32	48	69	83	100	118	138
SB60	16	32	48	69	86	102	118	132
SB61	12	28	44	64	83	100	114	132
SB62	12	28	44	61	83	100	114	132
SB63	13	29	45	65	85	101	116	134
SB64	14	30	46	66	83	98	114	132
SB65	16	32	48	69	83	100	118	132
SB66	15	31	47	68	81	94	119	132
SB67	17	33	48	69	86	102	119	138
SB68	17	33	49	69	86	102	118	138
SB69	10	26	47	68	86	102	119	138
SB70	16	32	49	70	86	102	118	138
SB71	12	28	48	66	86	102	114	132
SB72	10	25	48	69	86	102	119	138

First-order groups of subjects generated by the analysis of the primary school group results.

The subjects are listed in the order generated by the cluster analysis program, which corresponds approximately to decreasing order of weighting to the group.

Group 1: SB12, 29, 28, 43, 46, 56.
Group 2: SB22, 42, 62, 20, 23, 32, 33, 24, 30, 31, 9, 38, 40, 52, 61, 53, 44, 50, 18, 71, 41, 43, 36, 51, 58, 65, 21, 16.
Group 3: SB57, 59, 58, 65, 36, 56, 60, 35, 47, 45, 70, 67, 72, 68, 48, 55, 54.
Group 4: SB1, 3, 7, 4, 16, 50.
Group 5: SB5, 8, 25, 27, 37.
Group 6: SB69, 72, 67, 45, 47, 68, 60, 70, 48, 57, 59, 58, 71, 54.
Group 7: SB2, 25, 8, 27, 26.
Group 8: SB15, 17, 19, 43.
Group 9: SB23, 64, 32, 33, 53, 22, 42, 62, 24, 30, 31, 39, 38, 40, 52, 61, 20, 44, 18, 50, 71, 41, 36, 58, 65, 51, 21, 16, 27.
Group 10: SB34, 35, 56, 60, 65, 57, 59, 36, 58, 47, 70, 55, 54.
Group 11: SB46, 66, 69, 45, 48.
Group 12: SB1, 11, 3, 4, 16, 50.
Group 13: SB4, 6, 7, 3, 16.
Group 14: SB10, 13, 49, 2, 26.
Group 15: SB14, 26, 2, 49.
Group 16: SB4, 9, 7.

Clusters generated by the analysis of the primary school group results.

The groups are listed in the order generated by the analysis program, roughly corresponding to decreasing order of weighting to the cluster.

Cluster 1: Group 7, group 14, group 15, group 5.

Cluster 2: Group 13, group 16, group 4, group 12.

Cluster 3: Group 3, group 6, group 10, group 2, group 9, group 1, group 11, group 8.

CATEGORY NUMBERING SCHEME FOR ANALYSIS OF THE COMBINED PRIMARY AND SECONDARY SCHOOL GROUPS' RESULTS (FIRST BEADS EXPERIMENT).

PROBLEM B1				PROBLEM B2				PROBLEM B3				PROBLEM B4			
GREEN		RED		GREEN		RED		GREEN		RED		GREEN		RED	
CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER
FU	1	FU	16	FU, FRF	32	FU	46	FUL	65	FUL, FRFL	80	FU, FRF	97	FU, FRF, FQANT	113
FSR, FQS	2	FSR, FQS	17	FSR	33	FSR, FQS, FPR	47	FUM	66	FUM	81	FSR, FQS, FPR	98	FSR, FQS, FPR	114
PSR, PQS, PPR	3	PSR, PQS, PPR	18	PSR, PQS	34	PSR, PQS, PPR	48	FSRL, FQSL, FPRL	67	FSRL, FQSL	82	PSR, PQS	99	PSR, PQS	115
PU, PRF	4	PU	19	PU, PRF	35	PU, PRF	49	FSRM, FPRM	68	FSRM, FQSM, FPRM	83	PU	100	PU, PQRF	116
FQA	5	PQA	20	FQA	36	FQA	50	FQAM	69	PU	84	FQA	101	PLL	117
PQA	6	PQANT	21	PQA	37	FQTA	51	FLL	70	FQAL	85	PQA	102	FQTA	118
PQTA	7	PQTA	22	PQTA	38	FQTM	52	FMM	71	FQAM	86	PQTA	103	FQTM	119
PQTM	8	PQTM	23	PQTM	39	FQNTM	53	FQTL	72	FLL	87	PQTM	104	FQTS	120
PQTR	9	PQTS	24	PQTS	40	PQA	54	FQTA	73	FMM	88	PQTS	105	FQTR	121
PQNTL	10	PQTR	25	FQTL	41	PQTA	55	FQTM	74	FQTL	89	PQTR	106	PQNTL	122
PQRA, PQROA	11	PQNTL	26	PQRA	42	PQTM	56	FQTS	75	FQTA	90	FQNTL	107	PQTL	123
PQRM	12	PQRA, PQROA	27	PQRM	43	PQNTA	57	FQTR	76	FQTM	91	PQRA	108	FQRA	124
PQRS	13	PQRL, PQRM	28	PQRS	44	PQNTL	58	PQRA	77	FQTS	92	PQRM	109	FQRM	125
PQRR	14	PQRS	29	U	45	FQRM	59	PQRM	78	FQTR	93	PQRS	110	PQRA	126
U	15	PQRR	30			PQRA	60	U	79	PQRA	94	PQRR	111	PQRM	127
		U	31			PQRM	61			PQRM	95	U	112	PQRS	128
						PQRS	62			U	96			U	129
						PQRR	63								
						U	64								

DATA FOR CLUSTER ANALYSIS OF THE COMBINED RESULTS FROM THE PRIMARY AND SECONDARY SCHOOL SUBJECTS.

SUBJECT	CATEGORIES.							
SB1	1	16	35	49	65	81	97	113
SB2	3	18	33	47	66	82	98	114
SB3	1	16	35	49	66	81	100	116
SB4	4	19	35	49	66	80	97	113
SB5	2	17	34	48	67	83	99	116
SB6	4	18	35	49	66	80	98	114
SB7	4	19	35	49	65	81	100	116
SB8	3	18	34	48	67	83	99	114
SB9	4	19	32	48	67	83	97	113
SB10	2	17	33	47	68	82	98	114
SB11	1	16	32	46	74	91	97	114
SB12	7	22	38	51	73	90	103	118
SB13	5	16	33	47	67	85	99	114
SB14	3	18	39	48	74	82	98	115
SB15	8	23	36	50	69	85	101	113
SB16	4	16	35	49	65	91	104	116
SB17	6	23	37	55	69	85	101	119
SB18	8	23	41	46	74	91	104	116
SB19	6	20	37	51	69	86	102	118
SB20	8	23	39	52	74	91	104	123
SB21	6	20	37	54	74	91	102	119
SB22	8	23	39	52	74	91	104	119
SB23	8	23	39	58	74	91	104	119
SB24	8	23	39	61	74	91	104	119
SB25	3	18	33	47	67	83	99	115
SB26	3	18	34	48	67	83	98	114
SB27	2	18	34	47	70	87	104	119
SB28	7	22	39	51	73	90	103	118
SB29	7	22	38	55	73	90	103	118
SB30	8	23	39	55	74	91	104	119
SB31	8	23	39	55	74	91	104	119
SB32	8	23	39	58	74	91	104	119
SB33	8	23	39	58	74	91	104	119
SB34	12	28	39	60	71	87	101	124
SB35	12	28	43	61	71	87	109	117
SB36	12	28	43	61	74	91	104	119

(CONTINUED OVERLEAF)

DATA FOR COMBINED CLUSTER ANALYSIS (CONT'D).

SUBJECT	CATEGORIES						
SB37	3	18	34	47	68	83	104	116
SB38	8	23	39	56	74	91	104	119
SB39	8	23	39	55	74	91	104	119
SB40	8	23	39	56	74	91	104	119
SB41	8	23	39	52	75	91	105	120
SB42	8	23	39	52	74	91	104	119
SB43	8	23	38	51	73	90	104	119
SB44	8	23	39	58	78	95	104	119
SB45	11	27	43	61	78	95	109	127
SB46	11	27	42	60	73	90	108	118
SB47	12	28	43	61	78	95	108	127
SB48	11	27	39	58	78	95	104	127
SB49	2	18	34	47	68	83	98	114
SB50	8	23	39	49	72	91	104	116
SB51	1	16	39	56	67	91	104	123
SB52	8	23	39	56	74	91	104	119
SB53	8	23	39	58	74	81	104	119
SB54	10	28	39	58	78	95	104	126
SB55	12	28	41	59	66	89	109	127
SB56	12	28	43	61	73	90	109	119
SB57	12	28	43	61	74	91	109	127
SB58	12	28	43	61	74	91	104	127
SB59	12	28	43	61	74	91	109	127
SB60	12	28	43	61	78	91	109	119
SB61	8	23	39	56	74	91	104	119
SB62	8	23	39	52	74	91	104	119
SB63	9	25	40	57	76	93	106	121
SB64	10	26	41	58	74	89	104	119
SB65	12	28	43	61	74	91	109	119
SB66	11	27	42	60	72	84	110	119
SB67	13	29	43	61	78	95	110	127
SB68	13	29	44	61	78	95	109	127
SB69	6	21	42	60	78	95	110	127
SB70	12	28	44	62	78	95	109	127
SB71	8	23	43	58	78	95	104	119
SB72	6	20	43	61	78	95	110	127
SB73	8	23	39	51	73	90	104	119
SB74	8	23	39	55	74	91	104	119
SB75	8	23	39	52	74	91	104	119
SB76	8	23	39	58	74	91	104	119
SB77	8	23	39	61	74	91	104	119
SB78	12	29	43	61	74	91	109	119

DATA FOR COMBINED CLUSTER ANALYSIS (CONCLUDED).

SUBJECT	CATEGORIES							
SB79	12	28	43	61	74	91	109	127
SB80	12	28	44	62	78	95	110	119
SB81	12	28	43	62	65	88	110	125
SB82	12	28	43	61	78	95	109	127
SB83	12	28	43	61	78	95	109	127
SB84	12	28	43	61	78	95	109	127
SB85	8	23	39	53	74	91	104	119
SB86	8	23	39	52	74	91	104	119
SB87	8	24	39	52	75	92	104	120
SB88	8	23	40	52	74	91	104	120
SB89	8	23	39	58	74	80	104	128
SB90	11	27	43	58	75	92	104	128
SB91	12	28	43	61	74	91	109	119
SB92	12	28	43	61	74	91	109	118
SB93	12	28	43	61	70	88	110	127
SB94	13	28	39	60	78	95	110	128
SB95	8	23	43	61	74	91	107	127
SB96	12	28	43	61	78	95	109	127
SB97	15	31	45	64	79	96	112	129
SB98	8	23	39	48	74	91	104	119
SB99	8	23	39	52	71	88	104	119
SB100	8	23	39	58	71	91	104	120
SB101	12	28	39	62	72	89	104	127
SB102	12	28	43	62	72	89	109	127
SB103	12	28	43	61	72	89	110	127
SB104	12	28	43	61	70	87	109	127
SB105	12	28	44	61	73	91	109	127
SB106	12	28	39	60	74	91	109	127
SB107	13	29	42	60	78	95	110	128
SB108	14	30	43	63	78	95	111	127
SB109	8	23	38	51	74	91	103	119
SB110	7	22	38	51	73	90	104	118
SB111	8	23	39	61	74	91	104	127
SB112	12	28	43	61	74	91	109	119
SB113	12	28	43	61	70	87	109	122
SB114	12	28	43	61	78	95	109	119
SB115	12	30	43	61	78	95	107	127
SB116	12	28	43	61	78	95	109	127
SB117	12	28	44	60	77	94	109	128
SB118	12	28	43	61	78	95	109	127
SB119	12	28	43	61	78	95	109	127
SB120	12	28	43	61	78	95	109	127

First-order groups of subjects generated by the analysis of the combined primary and secondary school subjects' results.

The subjects are listed in the order printed out by the cluster analysis program, roughly corresponding to decreasing order of weighting to the group.

Group 1: SB22, 42, 62, 75, [illegible], [illegible], [illegible], [illegible], [illegible], 7[illegible], [illegible], 77, 30, [illegible], 39, 74, 38, 40, [illegible], 61, [illegible], 9[illegible], 5[illegible], 111, 88, 89, 100, [illegible], [illegible], 50, 18, 73, 109, [illegible], 71, [illegible], 95, [illegible], 87, 5[illegible], 51, 16.

Group 2: SB57, 59, 79, 53, 65, 91, 112, [illegible], 36, 60, 78, 114, 56, 82, 83, 84, 96, 116, 118, 119, 120, 47, 104, 105, 11[illegible], 35, 93, 103, 102, 115, 70, 45, 106, 55, 67, 72, 68, 95, 101, 10[illegible], 111.

Group 3: SB8, 26, 5, 25, 37, 49, 27, 10, 13.

Group 4: SB12, 28, 29, 110, 43, 73, 109, 56, 105.

Group 5: SB70, 80, 82, 83, 84, 96, 116, 118, 119, 120, 114, 47, 60, 57, 59, 79, 104, 65, 91, 112, 92, 5[illegible], 5[illegible], 105, 36, 78, 113, 35, 93, 10[illegible], 102, 115, 45, 10[illegible], 55, 67, 72, [illegible]8, 101, 108.

Group 6: SB9[illegible], 107, 67, 68, 69, 7[illegible], [illegible], 8[illegible], 84, 96, 116, 118, 119, 120, 47, 11[illegible], 45, 115, 70, [illegible], 9[illegible], 1[illegible], [illegible], 1[illegible], 106, 48, 54.

Group 7: SB1, 3, 7, 16, [illegible], 6, 59.

Group 8: SB2, 10, 49, [illegible], [illegible], [illegible], [illegible]7, [illegible], [illegible], 6.

Group 9: SB31, 93, 103, 102, 1[illegible], 57, 59, 79, 53, 65, 91, 112, 9[illegible], 36, 60, 78, 82, 83, 84, 96, 11[illegible], 118, 119, 120, 11[illegible], 47, 56, 113, 35, 105, 115, 70, 45, 106, 55, 67, 7[illegible], 80, 101, 108.

Group 10: SB4, 9, 1, 7, 16, 6.

Group 11: SB15, 17, [illegible]0, 31, [illegible], 7[illegible], 22, [illegible], [illegible], 75, 86, [illegible], 32, 33, 76, 24, 77, 38, 40, 52, 61, 85, 98, 20, 5[illegible], 111, [illegible], 89, 100, 99, 44, 50, 73, 18, 109, 71, [illegible], [illegible], 95.

Group 12: SB19, 21, 17, 109, [illegible], 7[illegible].

Group 13: SB[illegible], 64, 32, 33, 7[illegible], 5[illegible], [illegible], 4[illegible], 62, 75, [illegible], [illegible], 77, [illegible], [illegible], 29, 74, 38, 40, 52, 61, 85, 98, [illegible], 111, [illegible], [illegible], 10[illegible], 99, [illegible], 18, 50, 73, 109, 71, 43, 95, [illegible], 87, 5[illegible], [illegible], 16.

Group 14: SB[illegible], 35, 104, 113, 57, 59, 79, 6[illegible], 91, 11[illegible], [illegible], [illegible], [illegible], 5[illegible], 7[illegible], 114, 82, 83, 84, 96, 11[illegible], 11[illegible], [illegible], [illegible], [illegible], [illegible], 105, [illegible], 10[illegible], 102, 106, 70, 115, 5[illegible], [illegible], 81, [illegible], 117.

Group 15: SB[illegible]6, [illegible]6, 69, 107, [illegible], [illegible], 45, [illegible].

Group 16: SB[illegible], [illegible], [illegible], 7[illegible], [illegible], [illegible], 8[illegible], [illegible], [illegible], [illegible], [illegible], 11[illegible], 11[illegible], 1[illegible], 114, 47, 115, 60, 67, 7[illegible], [illegible], [illegible], [illegible], [illegible], [illegible], 111.

Group 17: SB[illegible], 1[illegible], [illegible], [illegible], [illegible], [illegible], [illegible].

Group 18: SB63, 83.

Clusters generated by the analysis of the combined primary and secondary School subjects' results.

The groups are listed in the order generated by the cluster analysis program, roughly corresponding to decreasing order of weighting to the cluster.

Cluster 1: Group 3, group [illegible], group [illegible], group 7, group 17.

Cluster 2: Group 11, group 18, group 12, group 1, group 13, group 17, group 7, group 8.

Cluster 3: Group 5, group 9, group 14, group 6, group 2, group 16, group 15, group 4, group 1, group 13.

APPENDIX D

Results of the control experiment carried out to validate the result of the first beads experiment with different materials.

Contents: Page:

TABLE OF INTERPRETATIVE CATEGORISATIONS OF RESPONSES BY SUBJECTS IN THE CONTROL EXPERIMENT WHO HAD PREVIOUSLY BEEN INVOLVED IN THE FIRST BEADS EXPERIMENT, TOGETHER WITH THEIR RESPONSES TO THE FIRST BEADS EXPERIMENT.

	SUBJECT	PROBLEM C1 (B1)		PROBLEM C2 (B2)		PROBLEM C3 (B3)		PROBLEM C4 (B4)		PROBLEM C5 (B5	
		GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
AGED 8 YRS ON BOTH OCCASIONS	SB43	2 (2)	2 (2)	2 (2)	2a (2a)	2 (2)	2 (2)	2 (2)	2 (2)	(2)	(2
	SB44	2 (2)	2 (2)	2 (2)	2b (2b)	2 (3)	2 (3)	2 (2)	2? (2)	2 (3)	3 (2)
	SB45	2 (2)	2 (2)	2 (3)	2a (3)	3 (3)	3 (3)	3? (3)	2 (3)		
AGED 8 WHEN TESTED FOR BEADS EXPERIMENT, 9 IN CONTROL EXPERIMENT	SB37	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (2)	1 (1)		
	SB46	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (2)	(3)	(3)
	SB47	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)	3 (3)		
	SB48	3 (3)	3 (3)	3 (2)	3 (2b)	3 (3)	3 (3)	3 (2)	3 (3)		
AGED 9 ON BOTH OCCASIONS	SB49	3 (1)	3 (1)	3 (1)	3 (1)	? (1)	? (1)	3 (1)	? (1)		
	SB53	2 (2)	2 (2)	2 (2)	2b (2b)	2 (2)	2 (2)	2 (2)	2 (2)		
	SB54	2 (3)	2 (2)	2 (2)	2b (2b)	2 (3)	2 (3)	2 (2)	2 (?)		
	SB57	3 (3)	3 (3)	3 (3)	3 (3)	3 (2)	3 (2)	3 (3)	3 (3)		
	SB58	3 (3)	3 (3)	3 (3)	3 (3)	2 (2)	2 (2)	3 (3)	3 (3)		

(FIGURES IN BRACKETS INDICATE RESPONSES TO FIRST BEADS EXPERIMENT ITEMS).

TABLE OF INTERPRETATIVE CATEGORISATIONS OF RESPONSES BY SUBJECTS WHO HAD NOT PREVIOUSLY BEEN INVOLVED IN THE FIRST BEADS EXPERIMENT.

SUBJECT	AGE (YEARS)	PROBLEM C1		PROBLEM C2		PROBLEM C3		PROBLEM C4		PROBLEM C5	
		GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SC1	8	1?	1?	1?	1?	2	1?	1?	1?		
SC2	8	2	2	2	2(a)	2	2	2	2	2	2
SC3	8	2	2	2	2(a)	2	2	2	2	2	2
SC4	8	2	2	2	2(a)	2	2	2	2		
SC5	8	?	3	?	?	?	2	2	?		
SC6	8	3	2	2	3	2	2	3	3		
SC7	9	2	2	2	2(a)	2	2	2	2		
SC8	9	3	3	?	?	2	?	?	?		
SC9	9	3	2	3	3	2	2	2	2		
SC10	9	3	3	3	3	2	2	3	2		
SC11	9	3	3	3	3	2	2	3	3		
SC12	9	3	3	3	3	3	3	3	2	3	2

APPENDIX D

Interpretative analysis of results of the roulette experiment.

Contents: Page:

TABLE OF INTERPRETATIVE CATEGORISATIONS OF SUBJECTS' RESPONSES TO THE ROULETTE EXPERIMENT.

SUBJECT	AGE (YEARS)	ORDER OF PRESENTATION OF PROBLEMS	PROBLEM R1		PROBLEM R2		PROBLEM R3		PROBLEM R4		PROBLEM R5	
			RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN
SR1	6	45312	1	1	1	1	1	1	1	1	1	1
SR2	6	53142	1	1	1	1	1	1	1	1	1	1
SR3	6	32451	1	1	1	1	1	1	1	1	1	1
SR4	6	24513	1	1	1	1	1	1	1	1	1	1
SR5	6	13254	1	1	1	1	1	1	1	1	1	1
SR6	6	52431	1	1	1	1	1	1	1	1	1	1
SR7	6	12534	1	1	1	1	1	1	1	1	1	1
SR8	6	[illegible]	1	1	1	1	1	1	1	1	1	1
SR9	6	32541	1	1	1	1	1	1	1	1	1	1
SR10	6	31524	1	1	1	1	1	1	1	1	1	1
SR11	6	52314	1	1	1	1	1	1	1	1	1	1
SR12	6	13542	1	1	1	3 ?	3 ?	3 ?	1	1	1	1
SR13	7	23541	1	1	1	1	1	1	1	1	1	1
SR14	7	[illegible]	1	1	1	1	1	1	1	1	1	1
SR15	7	23145	1	1	1	1	1	1	1	1	1	1
SR16	7	45[illegible]	1	1	1	1	1	1	1	1	1	1
SR17	7	12543	1	1	1	1	1	1	1	1	1	1
SR18	7	[illegible]	1	1	1	1	1	1	1	1	1	1
SR19	7	41253	1	1	1	1	1	1	1	1	1	1
SR20	7	25341	1	1	1	1	1	1	1	1	1	1
SR21	7	12543	1	1	1	1	1	2	1	1	1	1
SR22	7	52314	1	2	1	2	2	1	2	?	1	?
SR23	7	[illegible]	2	2	1	2	3	3	2	2	2	2
SR24	7	[illegible]34176	2	2 ?	3	3	3	3	3	3	2	2
SR25	8	[illegible]	1	1	1	1	1	1	1	1	1	1
SR26	8	41352	1	1	1	1	1	1	1	1	1	1
SR27	8	12534	1	1	1	1	1	1	1	1	1	1
SR28	8	41352	1	1	1	1	1	1	1	1	1	1
SR29	8	41235	1	1	1	1	1	1	1	1	1	1
SR30	8	41253	1	1	1	1	1	1	1	1	1	1
SR31	8	23514	1	1	1	2	2	2	2	1	1	1
SR32	8	13254	1	1	1	1	2	3	2	3	1	1
SR33	8	14532	1	1	3 ?	3	3	3	1	1	1	1
SR34	8	2153476	2	?	1	2	2	2	2	2	2	?
SR35	8	[illegible]	2	2	2	2	3	3	2	2	2	2
SR36	8	2354176	2	2	2	2	3	3	2	2	2	2

(CONTINUED OVERLEAF)

INTERPRETATIVE CATEGORISATIONS OF ROULETTE RESULTS (CONTINUED).

SUBJECT	AGE (YEARS)	ORDER OF PRESENTATION OF PROBLEMS	PROBLEM R1		PROBLEM R2		PROBLEM R3		PROBLEM R4		PROBLEM R5	
			RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN
SR37	9	25314	2	2	1	1	1	2	2	2	1	1
SR38	9	34521	1	2	?	1	3	3	2	2	1	1
SR39	9	25314	2	2	2	1	1	2	2	1	2	2
SR40	9	14532	1	1	2	2	2?	?	2	3	2	1
SR41	9	14325	1	2	1	1	2	2	2	2	2	2
SR42	9	52314	2	?	2	2	2	2	2	3	1	1
SR43	9	32514	2	2	1	2	2	2	2	2	2	2
SR44	9	43512	?	3	3	3	3	3	1	3	2	2
SR45	9	5231476	2	2	2	2	2	2	2	3	2	2
SR46	9	3214576	?	2	2	2	3	3	2	2	2	2
SR47	9	2345167	2	2	3?	3	3	3	3?	3	2	2
SR48	9	14325	4	4	2	2	2	2	2	2	3	3
SR49	10	34512	1	1	1	1	1	1	1	1	1	1
SR50	10	14532	2	1	2	2	2	2	1	1	2	2
SR51	10	3254176	4	4	2	1	1	1	2	2	2	2
SR52	10	13541	2	2	2	2	1	2	1	2	2	2
SR53	10	4325167	2	2	2	2	2	2	2	2	1	1
SR54	10	24315	2	3	2	2	2	2	2	?	1	1
SR55	10	12543	2	2	1	2	2	2	2	2	2	2
SR56	10	45312	2	2	2	2	2	3	2	2	1	2
SR57	10	521347	2	2	2	2	2	2	?	2	2	2
SR58	10	3254176	2	2	3	2	3	2	2	3	2	2
SR59	10	4125367	2	2	2	3	3	3	2	3	2	2
SR60	10	2354176	3	3	3	3	3	3	3	3	3	3
SR61	11	12345	1	1	1	2	1	1	1	1	1	1
SR62	11	41523	2	1	1	1	1	1	1	2	1	1
SR63	11	12543	2	3	2	3	2	3	1	2	2	1
SR64	11	34521	2	2	1	2	2	2	2	2	2	2
SR65	11	41235	2	2	2	2	3	2	2	3	2	1
SR66	11	1452376	4	4	2	2	1	2	3	3	4	4
SR67	11	25314	2	2	2	2	2	2	2	2	2	2
SR68	11	53241	2	2	2	2	2	3	2	2	2	2
SR69	11	2314567	?	3	2	2	2	2	2	2	2	2
SR70	11	34251	2	2	2	3	3	3	2	2	2	2
SR71	11	1253467	2	2	3	3	3	3	3	3	3	3
SR72	11	3421567	4	4	2	3	3	3	2	2	4	4

TABLE OF THE FREQUENCY OF THE DIFFERENT TYPES OF RESPONSE AT DIFFERENT AGES, USING RESULTS FROM PROBLEMS R1-R5, MULTIPLIED BY $\frac{4}{5}$ TO FACILITATE COMPARISON WITH THE OTHER FREQUENCY TABLES.

AGE GROUP	NUMBER OF TYPE 1 RESPONSES $\times \frac{4}{5}$	NUMBER OF TYPE 2 RESPONSES $\times \frac{4}{5}$	NUMBER OF TYPE 3 RESPONSES $\times \frac{4}{5}$	NUMBER OF TYPE 4 RESPONSES $\times \frac{4}{5}$	NUMBER OF UNCLASSIFIABLE RESPONSES $\times \frac{4}{5}$
6	93·6	0	2·4	0	0
7	75·2	12·8	6·4	0	1·6
8	63·2	23·2	8	0	1·6
9	17·6	56	16·8	1·6	4
10	19·2	58·4	15·2	1·6	1·6
11	17·6	52·8	18·4	6·4	0·8

TABLE OF THE FREQUENCY OF THE DIFFERENT TYPES OF RESPONSE AT DIFFERENT AGES, USING RESULTS FROM PROBLEMS R2-R5.

AGE GROUP	NUMBER OF TYPE 1 RESPONSES	NUMBER OF TYPE 2 RESPONSES	NUMBER OF TYPE 3 RESPONSES	NUMBER OF TYPE 4 RESPONSES	NUMBER OF UNCLASSIFIABLE RESPONSES
6	93	0	3	0	0
7	75	10	9	0	2
8	61	24	10	0	1
9	18	56	20	0	2
10	21	57	16	0	2
11	19	52	21	4	0

APPENDIX F

Cluster analysis of results of the roulette experiment.

Contents: | Page:

TABLE OF INITIAL CATEGORISATIONS OF SUBJECTS' RESPONSES TO THE ROULETTE EXPERIMENT ITEMS.

SUBJECT	AGE (YEARS)	PROBLEM R1		PROBLEM R2		PROBLEM R3		PROBLEM R4		PROBLEM R5	
		RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN
SR1	6	FSRS	FSRB	PSR	PSR	PSR	PSR	PSR	PSR	FSRS	FSRB
SR2	6	FSRS	FSRB	PSR	PSR	PSR	PSR	FSR	FSR	FUS	FSRB
SR3	6	FSRB	FSRS	FSR	FSR	FSR	FSR	PSR	PSR	FSRB	FSRS
SR4	6	FSRB	FSRS	FSR	FSR	FSR	FSR	PSR	PSR	FSRB	FSRS
SR5	6	FSRB	FSRS	FSR	FSR	FSR	FSR	PSR	PSR	FSRB	FSRS
SR6	6	FUB	FUS	FU	FU	FU	FU	PU	PU	FUS	FUB
SR7	6	FPRB	FPRB	PPR	PPR	PPR	FPR	FPR	FPR	FPRB	FPRS
SR8	6	FSRB	FSRS	PSR	PSR	FSR	FSR	PSR	PSR	FSRB	FSRS
SR9	6	FUB	FUS	FU	FU	PRF	PRF	FU	FRF	FRFB	FRFS
SR10	6	FSRB	FSRS	FSR	FSR	PSR	PSR	PSR	PSR	FSRB	FSRS
SR11	6	FUB	FPRS	FPR	PPR	PPR	PPR	FU	FPR	FSRB	FPRS
SR12	6	FSRB	FPRB	PQS	PQRO	PQRO	PQRO	FSR	FSR	FRFB	FSRS
SR13	7	FSRB	FSRB	FSR	FSR	FSR	FSR	FSR	FSR	FSRB	FSRS
SR14	7	FRFB	FRFB	PRF	PRF	FSR	FRF	PRF	PSR	FRFS	FRFB
SR15	7	FPRS	FPRB	PPR	FU	FPR	FPR	FPR	FPR	FPRS	FPRB
SR16	7	FSRS	FSRB	FSR	PSR	FSR	FSR	FSR	FSR	FSRS	FSRS
SR17	7	FSRS	FSRB	FSR	FSR	FSR	FSR	PSR	PSR	FSRS	FSRB
SR18	7	FUS	FUB	FPR	PPR	PU	PU	FU	FU	FUS	FUB
SR19	7	FSRB	FSRS	PSR	PSR	FSR	FSR	PSR	PSR	FSRS	FSRB
SR20	7	FSRS	FUB	FSR	PSR	PU	FU	FRF	FU	FUB	FSRB
SR21	7	FSRS	FSRB	FSR	FPR	PPR	PQNTE	FPR	FPR	FRFB	FSRB
SR22	7	FSRS	FQSB	FRF	PQTE	PQTE	PSR	PQTE	FRF	FSRB	FRFS
SR23	7	FQTE	FQTE	PU	PQTE	PQRE	PQRE	PQTE	FQTE	PQTF	PQTF
SR24	7	FQTE	FQTE	PQRE	PQRE	PQRE	PQRE	PQRE	PQRE	FQTE	FQTE
SR25	8	FSRB	FSRS	PSR	PSR	PSR	PSR	PSR	PSR	FSRB	FSRS
SR26	8	FUS	FUB	PPR	PU	FU	FPR	PU	FPR	FPRS	FUS
SR27	8	FUB	FUS	PU	PU	FU	FU	PU	PU	FUB	FUS
SR28	8	FSRB	FQSS	FRF	FU	FSR	FSR	FSR	FSR	FQSB	FSRS
SR29	8	FSRB	FSRS	PSR	PSR	FSR	FSR	PSR	PPR	FSRB	FSRS
SR30	8	FSRB	FPRS	PRF	PPR	PU	PPR	FRF	FSR	FSRB	FSRS
SR31	8	FQROB	FSRS	PSR	PQTE	PQTE	PQTE	PQTE	PSR	FSRB	FSRS
SR32	8	FSRS	FPRB	PPR	PPR	PQTF	PQRF	PQTF	PQTF	FUB	FPRS
SR33	8	FSRB	FSRS	PQRF	PQRE	PQRE	PQRE	FSR	FRF	FRFS	FRFB
SR34	8	FQTE	FQTES	PPR	PQTF	PQTF	PQTF	PQTF	PQTF	FQTE	FRFS
SR35	8	FQTE	FQTE	FQTE	PQTE	PQRE	PQRE	PQTE	FQTE	FQTE	FQTE
SR36	8	FQTE	FQTE	FQTE	PQTE	PQRE	PQRE	PQTE	FQTE	FQTE	FQTE

(CONTINUED OVERLEAF).

INITIAL CATEGORISATIONS OF ROULETTE RESULTS (CONT'D).

SUBJECT	AGE (YEARS)	PROBLEM R1		PROBLEM R2		PROBLEM R3		PROBLEM R4		PROBLEM R5	
		RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED	GREEN
SR37	9	FQTE	FQTE	FRF	FRF	PU	PQTE	PQTE	PQTE	FUB	FPRB
SR38	9	FSRS	FQTE	FRF	FU	PQRE	PQRE	PQTE	FQTE	FSRS	FPRB
SR39	9	FQTE	FQTE	FQTE	FU	FSR	PQTE	PQTE	PPR	FQTE	FQTE
SR40	9	FSRS	FSRB	PQTE	PQTE	FQTE	FSR	PQTE	PQRE	FQTE	FSRS
SR41	9	FPRS	FQTE	PPR	FPR	FQTE	PQTE	PQTE	FQTE	FQTF	FQTF
SR42	9	FQTE	FQROS	FQTE	PQTE	PQTE	PQTE	PQTE	PQRE	FSRS	FSRB
SR43	9	FQTE	FQTE	PU	PQTE	FQTE	PQTE	PQTE	PQTE	PQRV	PQRV
SR44	9	FQROS	FQROB	PQROE	PQRE	PQRE	PQRE	PU	PQRE	PQRV	PQRV
SR45	9	FQTE	FQTE	PQNTE	PQTE	PQNTE	PQNTE	PQNTE	PQRE	FQTE	FQTE
SR46	9	FQTES	FQTE	FQTE	PQTE	PQRE	PQRE	PQTE	FQTE	FQTE	FQTE
SR47	9	PQRV	PQRV	PQRF	PQREL	PQRE	PQRE	PQTF	PQRE	PQRV	PQRV
SR48	9	PQRV	PQRV	PQTE	PQTE	PQNTE	PQNTE	PQNTF	PQNTE	PQRF	PQRF
SR49	10	FUS	FSRB	PPR	FPR	PPR	PPR	PU	PU	FUB	FUB
SR50	10	FQTW	FUS	FQTW	PQTE	PQTE	PQTE	FSR	FSR	PQRV	PQRV
SR51	10	PQRV	PQRV	PQTF	FQTF	PSR	PSR	PQTE	PQTE	PQRV	PQRV
SR52	10	FQTE	FQTE	FQTE	PQTE	PRF	PQTE	FPR	FQTE	FQTE	FQTE
SR53	10	FQTE	FQTE	PQTF	PQTF	PQTF	PQTF	PQTE	PQTE	FSRB	FSRB
SR54	10	FQTE	FQROS	FQTE	PQTE	PQTE	PQTE	PQTE	PRF	FSRB	FSRS
SR55	10	FQTE	FQTE	PRF	PQTE	PQTE	PQTE	PQTE	PQTF	FQTF	FQTE
SR56	10	FQTE	FQTE	PQNTF	PQTF	PQTE	PQRE	PQTE	FQTE	FUS	FSRB
SR57	10	FQTE	FQTE	PQNTE	PQTE	PQNTE	PQTE	FSR	FQTE	FQTE	FQTE
SR58	10	FQTE	FQTE	PQROE	PQTE	PQRE	PQTE	PQTE	PQRE	FQTE	FQTE
SR59	10	PQRV	PQRV	PQTF	PQRE	PQRE	PQRE	PQTF	PQRE	PQRV	PQRV
SR60	10	PQRE	PQRE	PQRF	PQRE	PQRE	PQRE	PQRF	PQRE	PQRF	PQRF
SR61	11	FUS	FPRS	PRF	PQTE	FPR	FPR	FU	PPR	FUB	FPRS
SR62	11	FQTW	FSRS	FU	PU	PSR	FSR	FSR	PQTE	FUB	FSRS
SR63	11	FQROB	FQROS	PQTF	PQRF	PQTE	PQRE	FU	FQTE	FQTE	FSRS
SR64	11	FQTE	FQTE	FU	PQTE	PQTE	PQTE	PQTE	PQTE	FQTE	FQTE
SR65	11	FQTE	FQTE	PQTF	PQTE	PQRE	PQTE	PQTF	PQRE	FQTE	FSRS
SR66	11	PQRV	PQRV	PQTE	PQTE	PU	PQTE	PQRF	PQRE	PQRV	PQRV
SR67	11	PQRV	PQRV	PQTE	PQTE	PQTE	PQTE	PQTE	PQTE	PQTF	PQTF
SR68	11	PQRF	PQRF	PQTE	PQTE	PQRF	PQTE	PQTE	PQTE	PQTF	PQTF
SR69	11	FSRS	FQRES	PQTF	PQTF	PQTF	PQTE	PQTF	PQTF	PQTF	PQTF
SR70	11	PQRV	PQRV	PQTF	PQRE	PQRF	PQRE	PQTF	PQTF	PQRV	PQRV
SR71	11	PQRV	PQRV	PQNTF	PQRE	PQRE	PQRE	PQTF	PQTE	PQRE	PQRE
SR72	11	PQRV	PQRV	PQTE	PQRF	PQRE	PQRE	PQTE	PQTF	PQRV	PQRV

CATEGORY NUMBERING SCHEME FOR ANALYSIS OF RESULTS FROM PROBLEMS R2 - R5 (ANALYSIS A).

PROBLEM R2				PROBLEM R3				PROBLEM R4				PROBLEM R5			
RED		GREEN		RED		GREEN		RED		GREEN		RED		GREEN	
CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER
PU	1	PU	19	PU	34	PU	48	PU	62	PU	75	FUB	88	FUB	103
PSR	2	PSR	20	PSR	35	PSR	49	PSR	63	PSR	76	FSRB	89	FSRB	104
PPR	3	PPR	21	PPR	36	PPR	50	FU	64	PPR	77	FPRB	90	FPRB	105
FU	4	FU	22	FU	37	FU	51	FSR	65	FU	78	FUS	91	FUS	106
FSR	5	FSR	23	FSR	38	FSR	52	FPR	66	FSR	79	FSRS	92	FSRS	107
FPR	6	FPR	24	FPR	39	FPR	53	PRF	67	FPR	80	FPRS	93	FPRS	108
PQS	7	PRF	25	PRF	40	PRF	54	FRF	68	PRF	81	FQSB	94	FRFB	109
PRF	8	FRF	26	FQTE	41	FRF	55	PQTE	69	FRF	82	FRFB	95	FRFS	110
FRF	9	FQTF	27	PQTE	42	PQTE	56	PQTF	70	FQTE	83	FRFS	96	FQTE	111
FQTW	10	PQTE	28	PQTF	43	PQTF	57	PQRE	71	PQTE	84	FQTE	97	FQTF	112
FQTE	11	PQTF	29	PQRE	44	PQRE	58	PQRF	72	PQTF	85	FQTF	98	PQTF	113
PQTE	12	PQRE	30	PQRF	45	PQRF	59	PQNTE	73	PQRE	86	PQTF	99	PQRE	114
PQTF	13	PQRF	31	PQNTE	46	PQNTE	60	PQNTF	74	PQNTE	87	PQRE	100	PQRF	115
PQRE	14	PQREL	32	PQRO	47	PQRO	61					PQRF	101	PQRV	116
PQRF	15	PQRO	33									PQRV	102		
PQNTE	16														
PQNTF	17														
PQROE	18														

DATA FOR CLUSTER ANALYSIS OF RESULTS FROM PROBLEMS R2-R5.

SUBJECT	CATEGORIES.						
SR1	2	20	35	49	63	76	92	104
SR2	2	20	35	49	65	79	91	104
SR3	5	23	38	52	63	76	89	107
SR4	5	23	38	52	63	76	89	107
SR5	5	23	38	52	63	76	89	107
SR6	4	22	37	51	62	75	91	103
SR7	3	21	36	53	66	80	90	108
SR8	2	20	38	52	63	76	89	107
SR9	4	22	40	54	64	82	95	110
SR10	5	23	35	49	63	76	89	107
SR11	6	21	36	50	64	80	89	108
SR12	7	33	47	61	65	79	95	107
SR13	5	23	38	52	65	79	89	107
SR14	8	25	38	55	67	76	96	109
SR15	3	22	39	53	66	80	93	109
SR16	5	20	38	52	65	79	92	107
SR17	5	23	38	52	63	76	92	104
SR18	6	21	34	48	64	78	91	103
SR19	2	20	38	52	63	76	92	104
SR20	5	20	34	51	68	78	88	104
SR21	5	24	36	60	66	80	95	104
SR22	9	28	42	49	69	82	89	110
SR23	1	28	44	58	69	83	99	113
SR24	14	30	44	58	71	86	97	111
SR25	2	20	35	49	63	76	89	107
SR26	3	19	37	53	62	80	93	106
SR27	1	19	37	51	62	75	88	106
SR28	9	22	38	52	65	79	94	107
SR29	2	20	38	52	63	77	89	107
SR30	8	21	34	50	68	79	89	107
SR31	2	28	42	56	69	76	89	107
SR32	3	21	43	59	70	85	88	108
SR33	15	30	44	58	65	82	96	109
SR34	3	29	43	57	70	85	97	110
SR35	11	28	44	58	69	83	97	111
SR36	11	28	44	58	69	83	97	111

(CONTINUED OVERLEAF)

DATA FOR CLUSTER ANALYSIS OF RESULTS FROM PROBLEMS R2-R5 (CONTINUED).

SUBJECT	CATEGORIES							
SR37	9	26	34	56	69	84	88	105
SR38	9	22	44	58	69	83	92	105
SR39	11	22	38	56	69	77	97	111
SR40	12	28	41	52	69	86	97	107
SR41	3	24	41	56	69	83	98	112
SR42	11	28	42	56	69	86	92	104
SR43	1	28	41	56	69	84	102	116
SR44	18	30	44	58	62	86	102	116
SR45	16	28	46	60	73	86	97	111
SR46	11	28	44	58	69	83	97	111
SR47	15	32	44	58	70	86	102	116
SR48	12	28	46	60	74	87	101	115
SR49	3	24	36	50	62	75	88	103
SR50	10	28	42	56	65	79	102	116
SR51	13	27	35	49	69	84	102	116
SR52	11	28	40	56	66	83	97	111
SR53	13	29	43	57	69	84	89	104
SR54	11	28	42	56	69	81	89	107
SR55	8	28	42	56	69	85	98	111
SR56	17	29	42	58	69	83	91	104
SR57	16	28	46	56	65	83	97	111
SR58	18	28	44	56	69	86	97	111
SR59	13	30	44	58	70	86	102	116
SR60	15	30	44	58	72	86	101	115
SR61	8	28	39	53	64	77	88	108
SR62	4	19	35	52	65	84	88	107
SR63	13	31	42	58	64	83	97	107
SR64	4	28	42	56	69	84	97	111
SR65	13	28	44	56	70	86	97	107
SR66	12	28	34	56	72	86	102	116
SR67	12	28	42	56	69	84	99	113
SR68	12	28	45	56	69	84	99	113
SR69	13	29	43	56	70	85	99	113
SR70	13	30	45	58	70	85	102	116
SR71	17	30	44	58	70	84	100	114
SR72	12	31	44	58	69	85	102	116

<u>First-order groups of subjects generated by the analysis of the results from problems R2 - R5.</u>

The subjects are listed in the order printed out by the analysis program, roughly corresponding to decreasing order of weighting to the group.

Group 1: SR36, 46, 35, 23, 58, 52, 57, 64, 39, 55, 42, 54, 67, 31, 68, 65, 40, 43, 45, 22, 50, 66, 61.

Group 2: SR4, 5, 3, 8, 29, 13, 10, 17, 19, 25, 16, 28, 31, 62, 49.

Group 3: SR44, 59, 47, 70, 72, 60, 24, 33, 71, 35, 36, 46, 23, 38, 65, 63.

Group 4: SR1, 25, 8, 19, 29, 4, 5, 3, 10, 17, 16, 31, 62.

Group 5: SR1, 2, 25, 19, 8, 29, 10, 3, 17, 16, 31, 62.

Group 6: SR24, 58, 35, 36, 46, 52, 57, 64, 39, 45, 65, 55, 42, 40, 23, 63.

Group 7: SR43, 51, 72, 59, 70, 44, 47, 66, 50, 65, 35, 36, 46, 23.

Group 8: SR34, 69, 32, 53, 70, 59, 65, 71, 72.

Group 9: SR7, 11, 15, 26, 21, 49.

Group 10: SR6, 27, 26, 49, 62.

Group 11: SR11, 30, 18, 7, 32, 61, 49.

Group 12: SR46, 56, 35, 36, 23, 38, 58, 52, 64, 57, 39, 42, 55, 54, 67, 31, 68, 40, 43, 22, 41, 37.

Group 13: SR44, 58, 59, 47, 24, 60, 65, 35, 36, 46, 23, 72, 38, 33, 71, 63.

Group 14: SR56, 63, 35, 36, 46, 23, 38, 58, 52, 64, 57, 39, 42, 55, 54, 31, 67, 40, 22, 41.

Group 15: SR67, 69, 68, 23, 64, 55, 58, 42, 54, 31, 43, 65, 52, 57, 39, 50, 66, 37, 41.

Group 16: SR53, 56, 42, 54, 31, 64, 67, 55, 68, 43, 58, 22, 35, 36, 46, 23, 40, 39, 33, 41, 37, 72, 51.

Group 17: SR12, 13, 16, 28, 4, 5, 3, 8, 29, 10, 25, 31, 62, 54, 40, 30, 65, 63.

Group 18: SR45, 48, 57, 52, 35, 36, 46, 58, 64, 55, 42, 54, 65, 40, 31, 67, 68, 23, 43, 22, 66, 50, 61.

Group 19: SR20, 30, 18, 27, 66, 61.

Group 20: SR14, 32, 13, 16, 28.

Clusters generated by the analysis of the results from problems R2 - R5

The groups are listed in the order printed out by the analysis program, roughly corresponding to decreasing order of weighting to the cluster.

Cluster 1:

Groups of cluster 1: group 9, group 11, group 10, group 19.

Subjects of cluster 1: SR7, 11, 18, 27, [illegible], 49, 61.

Subjects not of cluster 1: SR1, 3, 4, 5, [illegible], 10, [illegible], 14, 17, 19, 22, 24, 25, [illegible], 35, 36, [illegible], 45, 46, 47, 48, 56, 59, 60, 69, 70, 71.

Cluster 2:

Groups of cluster 2: group 2, group 4, group 5, group 20 group 17.

Subjects of cluster 2: SR1, 2, 3, 4, 5, 8, 10, 12, 13, 14, 16, 17, 19, 20, 25, 28, 29, 30, 31, 39, 40, 54, 62, 63, 65.

Subjects not of cluster 2: SR6, 7, 9, 18, 26, 27, [illegible], 34, 41, 45, 48, 49, 61, 69, 70, 72.

Cluster 3:

Groups of cluster 3: group 18, group 1[illegible], group 6, group 7, group 13, group 3, group 8, group 1, group 16, group 12, group 15, group [illegible].

Subjects of cluster 3: SR22, [illegible], 37, [illegible], [illegible], [illegible], [illegible], 50, 51, 52, 53, 54, 55, 57, 59, 6[illegible], [illegible]5, 66, 67, 68, 72.

Subjects not of cluster 3: SR1, 2, 3, 4, 5, 6, [illegible], 9, 10, 12, [illegible], 14, 15, 16, 17, 19, 21, 25, 26, 28.

Cluster 4:

Groups of cluster 4: group 17, group 19, group 2, group [illegible].

Subjects of cluster 4: SR11, 2[illegible], [illegible]2, 29, [illegible], 31, 39, [illegible], 42, 5[illegible], 54, 62, 65.

Subjects not of cluster 4: SR6, 9, 15, 2[illegible], [illegible], [illegible], [illegible], 38, 4[illegible], [illegible], 45, 47, 48, 5[illegible], 59, [illegible], 69, 70, 71.

Cluster 5:

Groups of cluster 5: group 15, group 17, group [illegible].

Subjects of cluster 5: SR22, [illegible], [illegible]7, 39, [illegible], [illegible], [illegible], 50, 51, 52, 5[illegible], 54, 55, 57, [illegible], [illegible], [illegible], [illegible]5, [illegible], [illegible]7, 68.

Subjects not of cluster 5: SR[illegible], [illegible], 15, [illegible], [illegible], 33, [illegible].

CATEGORY NUMBERING SCHEME FOR THE ANALYSIS OF RESULTS FROM PROBLEMS R1 – R5 (ANALYSIS B).

PROBLEM R1				PROBLEM R2				PROBLEM R3				PROBLEM R4				PROBLEM R5			
RED		GREEN		RED		GREEN		RED		GREEN		RED		GREEN		RED		GREEN	
CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER
FUB	1	FUB	15	PU	32	PU	47	PU	60	PU	73	PU	86	PU	99	FUB	112	FUB	127
FSRB	2	FSRB	16	PSR, PQS	33	PSR	48	PSR	61	PSR	74	PSR	87	PSR	100	FSRB	113	FSRB	128
FPRB	3	FPRB	17	PPR	34	PPR	49	PPR	62	PPR	75	FU	88	PPR	101	FPRB	114	FPRB	129
FUS	4	FUS	18	FU	35	FU	50	FU	63	FU	76	FSR	89	FU	102	FUS	115	FUS	130
FSRS	5	FSRS	19	FSR	36	FSR	51	FSR	64	FSR	77	FPR	90	FSR	103	FSRS	116	FSRS	131
FPRS	6	FPRS	20	FPR	37	FPR	52	FPR	65	FPR	78	PRF	91	FPR	104	FPRS	117	FPRS	132
FRFB	7	FQSB	21	PRF	38	PRF	53	PRF	66	PRF	79	FRF	92	PRF	105	FQSB	118	FRFB	133
FQTW, FQTE	8	FQSS	22	FRF	39	FRF	54	FQTE	67	FRF	80	PQTE	93	FRF	106	FRFB	119	FRFS	134
FQTES	9	FRFB	23	FQTW, FQTE	40	FQTF	55	PQTE	68	PQTE	81	PQTF	94	FQTE	107	FRFS	120	FQTE	135
PQRE	10	FQTE	24	PQTE	41	PQTE	56	PQTF	69	PQTF	82	PQRE	95	PQTE	108	FQTE	121	FQTF	136
PQRF	11	FQTES	25	PQTF	42	PQTF	57	PQRE, PQRO	70	PQRE, PQRO	83	PQRF	96	PQTF	109	FQTF	122	PQTF	137
PQRV	12	PQRE	26	PQRE, PQROE	43	PQRE, PQREL, PQRO	58	PQRF	71	PQRF	84	PQNTE	97	PQRE	110	PQTF	123	PQRE	138
FQROB	13	PQRF	27	PQRF	44	PQRF	59	PQNTE	72	PQNTE	85	PQNTF	98	PQNTE	111	PQRE	124	PQRF	139
FQROS	14	PQRV	28	PQNTE	45											PQRF	125	PQRV	140
		FQRVS	29	PQNTF	46											PQRV	126		
		FQROB	30																
		FQROS	31																

DATA FOR CLUSTER ANALYSIS OF RESULTS FROM PROBLEMS R1-R5.

SUBJECT	CATEGORIES								
SR1	5	16	33	48	61	74	87	100	116	128
SR2	5	16	33	48	61	74	89	103	115	128
SR3	2	19	36	51	64	77	87	100	113	131
SR4	2	19	36	51	64	77	87	100	113	131
SR5	2	19	36	51	64	77	87	100	113	131
SR6	1	18	35	50	63	76	86	99	115	127
SR7	3	17	34	49	62	78	90	104	114	132
SR8	2	19	33	48	64	77	87	100	113	131
SR9	1	18	35	50	66	79	88	106	119	134
SR10	2	19	36	51	61	74	87	100	113	131
SR11	1	20	37	49	62	75	88	104	113	132
SR12	2	17	33	58	70	83	89	103	119	131
SR13	2	16	36	51	64	77	89	103	113	131
SR14	7	23	38	53	64	80	91	100	120	133
SR15	6	17	34	50	65	78	90	104	117	129
SR16	5	16	36	48	64	77	89	103	116	131
SR17	5	16	36	51	64	77	87	100	116	128
SR18	4	15	37	49	60	73	88	102	115	127
SR19	2	19	33	48	64	77	87	100	116	128
SR20	5	15	36	48	60	76	92	102	112	128
SR21	5	16	36	52	62	85	90	104	119	128
SR22	5	21	39	56	68	74	93	106	113	134
SR23	8	24	32	56	70	83	93	107	123	137
SR24	8	24	43	58	70	83	95	110	121	135
SR25	2	19	33	48	61	74	87	100	113	131
SR26	4	15	34	47	63	78	86	104	117	130
SR27	1	18	32	47	63	76	86	99	112	130
SR28	2	22	39	50	64	77	89	103	118	131
SR29	2	19	33	48	64	77	87	101	113	131
SR30	2	20	38	49	60	75	92	103	113	131
SR31	13	19	33	56	68	81	93	100	113	131
SR32	5	17	34	49	69	84	94	109	112	132
SR33	2	19	44	58	70	83	89	106	120	133
SR34	8	25	34	57	69	82	94	109	121	134
SR35	8	24	40	56	70	83	93	107	121	135
SR36	8	24	40	56	70	83	93	107	121	135

(CONTINUED OVERLEAF)

DATA FOR CLUSTER ANALYSIS OF RESULTS FROM R1-R5 (CONT'D).

SUBJECT	CATEGORIES								
SR37	8	24	39	54	60	81	93	108	112	129
SR38	5	24	39	50	70	83	93	107	116	129
SR39	8	24	40	50	64	81	93	101	121	135
SR40	5	16	41	56	67	77	93	110	121	131
SR41	6	24	34	52	67	81	93	107	122	136
SR42	8	31	40	56	68	81	93	110	116	128
SR43	8	24	32	56	67	81	93	108	126	140
SR44	14	30	43	58	70	83	86	110	126	140
SR45	8	24	45	56	72	85	97	110	121	135
SR46	9	24	40	56	70	83	93	107	121	135
SR47	12	28	44	58	70	83	94	110	126	140
SR48	12	28	41	56	72	85	98	111	125	139
SR49	4	16	34	52	62	75	86	99	112	127
SR50	8	18	40	56	68	81	89	103	126	140
SR51	12	28	42	55	61	74	93	108	126	140
SR52	8	24	40	56	66	81	90	107	121	135
SR53	8	24	42	57	69	82	93	108	113	128
SR54	8	31	40	56	68	81	93	105	113	131
SR55	8	24	38	56	68	81	93	109	122	135
SR56	8	24	46	57	68	83	93	107	115	128
SR57	8	24	45	56	72	81	89	107	121	135
SR58	8	24	43	56	70	81	93	110	121	135
SR59	12	28	42	58	70	83	94	110	126	140
SR60	10	26	44	58	70	83	96	110	125	139
SR61	4	20	38	56	65	78	88	101	112	132
SR62	8	19	35	47	61	77	89	108	112	131
SR63	13	31	42	59	68	83	88	107	121	131
SR64	8	24	35	56	68	81	93	108	121	135
SR65	8	24	42	56	70	81	94	110	121	131
SR66	12	28	41	56	60	81	96	110	126	140
SR67	12	28	41	56	68	81	93	108	123	137
SR68	11	27	41	56	71	81	93	108	123	137
SR69	5	29	42	57	69	81	94	109	123	137
SR70	12	28	42	58	71	83	94	109	126	140
SR71	12	28	46	58	70	83	94	108	124	138
SR72	12	28	41	59	70	83	93	109	126	140

First-order groups of subjects generated by the analysis of the results from problems 21 - 25.

The subjects are listed in the order generated by the cluster analysis program, roughly corresponding to decreasing order of weighting to the group.

Group 1: SR25, 36, 46, 23, 5[illegible], 5[illegible], 5[illegible], [illegible], 45, 55, 6[illegible], [illegible], [illegible], 4[illegible], 54, 56, 50, 37, 53, 4[illegible], [illegible], [illegible].

Group 2: SR[illegible], 5, 3, 8, 1[illegible], [illegible], 1[illegible], 12, 17, 1[illegible], 1, [illegible]1, 6[illegible], [illegible].

Group 3: SR[illegible]7, 59, 70, 7[illegible], [illegible], 71, 51, [illegible], [illegible], 6[illegible], [illegible], [illegible]5, 3[illegible], 46.

Group 4: SR67, 68, 6[illegible], 43, 55, 59, 23, [illegible]5, 3[illegible], [illegible], 5[illegible], [illegible], [illegible], [illegible]5, 65, 42, 5[illegible], 50, 37, 5[illegible], 5[illegible], [illegible]1, [illegible], [illegible], 22.

Group 5: SR1, [illegible], 25, 1[illegible], 8, [illegible], 1[illegible], 17, 16, 1[illegible], 31, 6[illegible], 40.

Group 6: SR13, 2[illegible], 16, 3, 4, 5, [illegible], 2[illegible], [illegible], [illegible]5, 19, 17, [illegible]1, 6[illegible], [illegible].

Group 7: SR6, 27, 49, 28, 11.

Group 8: SR26, [illegible], 35, [illegible], [illegible], [illegible], 59, 57, [illegible], 3[illegible], 55, [illegible]5, [illegible], 65, 43, 4[illegible], 5[illegible], 56, 37, 5[illegible], 41, [illegible].

Group 9: SR66, 67, 68, [illegible], 6, 55, 5[illegible], [illegible], 5[illegible], 35, [illegible], 59, 4[illegible], [illegible], 57, [illegible], 6[illegible], [illegible], 5[illegible], [illegible]7, [illegible], [illegible], [illegible], [illegible].

Group 10: SR1[illegible], [illegible], [illegible]7, 5[illegible], [illegible], 70, 7[illegible], 7[illegible], 6[illegible], [illegible], 65, [illegible]5, [illegible], [illegible], [illegible], 3[illegible], 5[illegible], [illegible].

Group 11: SR52, 67, 7[illegible], [illegible], [illegible], [illegible]7, 7[illegible], 71, [illegible], [illegible], [illegible], [illegible], [illegible], 3[illegible], [illegible].

Group 12: [illegible]27, 15, [illegible], [illegible], 11, [illegible], [illegible].

Group 13: SR[illegible], 1[illegible], [illegible], 1[illegible], [illegible], [illegible], 5, [illegible], [illegible], 1[illegible], [illegible]5, 19, 3[illegible], 39, 6[illegible], 32.

Group 14: SR17, [illegible]1, 1, 16, [illegible], [illegible], [illegible], [illegible], [illegible], [illegible].

Group 15: SR[illegible], 69, 3[illegible], 52, 65, 55, 37, [illegible], [illegible], [illegible], 7[illegible].

Group 16: SR11, 30, 1[illegible], 61, 7, [illegible], [illegible].

Group 17: SR[illegible], 9, [illegible]7, 6[illegible], 5[illegible].

Group 18: SR51, 5[illegible], [illegible], 6[illegible], [illegible]7, [illegible], 5[illegible], [illegible], 35, [illegible], [illegible], [illegible], 65, [illegible], 54, 56, 67, [illegible], [illegible], [illegible], [illegible], [illegible], [illegible], 7[illegible].

Group 19: SR62, [illegible], [illegible], 37, 5[illegible], 5[illegible], [illegible], [illegible], [illegible], 5[illegible], 57, 45, 65, [illegible], [illegible], [illegible], 5[illegible], 56, 50, 5[illegible], [illegible], 3[illegible].

Group 20: SR5[illegible], 6[illegible], [illegible], 4[illegible], 64, [illegible], 58, 35, [illegible], [illegible], [illegible], 3[illegible], 5[illegible], [illegible], 45, 2[illegible], 56, 50, 67, 5[illegible], [illegible], [illegible], [illegible].

Group 21: SR2[illegible], 30, 23, [illegible]5, [illegible], [illegible], [illegible], 5[illegible], 6[illegible], [illegible]7, 2[illegible], 55, 45, [illegible], [illegible], [illegible], 5[illegible], 56, 37, 53, 4[illegible], 67, [illegible], [illegible], [illegible].

Group 22: SR67, 69, 6[illegible], [illegible], [illegible], [illegible], [illegible], 5[illegible], [illegible], [illegible], [illegible]7, 3[illegible], [illegible], [illegible], 5[illegible], 37, [illegible], 5[illegible], [illegible], [illegible], [illegible], [illegible].

<u>[illegible] groups of subjects [illegible] the [illegible] of the results from tables 21 - 35</u> (Cont.)

Group 23: [illegible], [illegible], 1, 25, 8, 12, 2[illegible], [illegible], [illegible], 16, [illegible], 5[illegible], [illegible].

Group 24: [illegible], 7[illegible], 5[illegible], [illegible], 7[illegible], 71, 5[illegible], [illegible], [illegible], [illegible], [illegible], 2[illegible].

Group 25: [illegible], [illegible], 17, 1, 1[illegible], [illegible], 1[illegible], [illegible], [illegible].

Group 26: [illegible], 20, 2, 2[illegible], [illegible], [illegible].

Group 27: [illegible], [illegible], 15, [illegible]6, [illegible], [illegible], 11.

Clusters [illegible] by the [illegible] of the results from problems 21 - [illegible].

The groups are listed in the order [illegible] by the [illegible] program, roughly corresponding to [illegible] order of weighting to the cluster.

Cluster 1:
Groups of cluster 1: group [illegible], group 16, group 7, group [illegible]7, group [illegible].
Subjects of cluster 1: SR6, 7, [illegible], [illegible], [illegible], [illegible], [illegible], [illegible]7, [illegible], [illegible], [illegible], 61.
Subjects not of cluster 1: SR[illegible], [illegible], 5, [illegible], [illegible], [illegible], [illegible], [illegible], 35, [illegible]6, [illegible],
39, [illegible], [illegible], [illegible], [illegible], [illegible]7, [illegible], [illegible], [illegible], [illegible], [illegible],
54, 55, 56, 57, 58, 59, 60, [illegible], [illegible], 65, 66,
67, 68, 70, [illegible], 7[illegible].

Cluster 2:
Groups of cluster 2: group 2, group 6, group [illegible], group 5, group 23,
group 1[illegible], group [illegible], group 26.
Subjects of cluster 2: SR1, [illegible], [illegible], 1[illegible], [illegible], 1[illegible], 1[illegible], 17, 19, 2[illegible], [illegible], 25,
28, [illegible], [illegible], [illegible], [illegible].
Subjects not of cluster 2. SR[illegible], 15, [illegible], [illegible], [illegible]5, 36, [illegible]1, [illegible], [illegible], [illegible]5, 46,
47, [illegible], 5[illegible], 5[illegible], [illegible], [illegible], 59, [illegible], [illegible], [illegible]7, [illegible],
7[illegible], 71, 7[illegible].

Cluster 3:
Groups of cluster 3: group [illegible], group [illegible], group [illegible], group [illegible], group [illegible],
group [illegible], group 1, group [illegible], group 1[illegible], group 3,
group 10, group 11, group [illegible], group 15, group 17.
Subjects of cluster 3: SR22, [illegible], 31, 35, 36, 37, [illegible], [illegible], [illegible], 50, 51, 5[illegible],
54, 55, 56, 57, [illegible], 6[illegible], 6[illegible], 65, [illegible], 67.
Subjects not of cluster 3: SR1, [illegible], [illegible], 5, 7, [illegible], 1[illegible], [illegible], 15, 17, 18, 1[illegible],
[illegible]1, [illegible]5, 2[illegible], [illegible].

Cluster 4:
Groups of cluster 4: group 7, group 17, group [illegible].
Subjects of cluster 4: SR6, [illegible], [illegible], [illegible], [illegible]7, [illegible], [illegible].
Subjects not of cluster 4: SR1, [illegible], [illegible], 5, [illegible], [illegible], [illegible], 17, 19, 2[illegible], [illegible]5, [illegible],
[illegible], 4[illegible], 4[illegible], [illegible], [illegible], [illegible]7, [illegible], 5[illegible], 5[illegible], 6[illegible], 6[illegible],
6[illegible], [illegible]9, 7[illegible], [illegible], 7[illegible].

Cluster 5:
Groups of cluster 5: group [illegible]5, group [illegible]7, group [illegible], group [illegible].
Subjects of cluster 5: SR1, [illegible], 1[illegible], 17, [illegible], [illegible], [illegible], [illegible], [illegible].
Subjects not of cluster 5: S[illegible], [illegible], [illegible], [illegible], [illegible], [illegible], [illegible], [illegible], [illegible], [illegible], [illegible]7,
48, 5[illegible], [illegible], [illegible], [illegible]7, 5[illegible], [illegible], [illegible], [illegible], [illegible], [illegible], [illegible],
67, [illegible], 7[illegible], 71, 7[illegible].

APPENDIX G

Interpretative analysis of results of the modified roulette experiment.

Contents: Page:

TABLE OF INTERPRETATIVE CATEGORISATIONS OF SUBJECTS' RESPONSES TO THE MODIFIED ROULETTE EXPERIMENT.

SUBJECT	AGE (YEARS)	ORDER OF PRESENT-ATION OF PROBLEMS	PROBLEM MR1		PROBLEM MR2		PROBLEM MR3		PROBLEM MR4	
			GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SMR1	6	2143	1	1	1	1	1	1	1	1
SMR2	6	3241	1	1	1	1	1	1	1	1
SMR3	6	1243	1	1	1	1	1	1	1	1
SMR4	6	3124	1	1	1	1	1	1	1	1
SMR5	6	4132	1	1	1	1	1	1	1	1
SMR6	6	1423	1	1	1	1	1	1	1	1
SMR7	6	3241	1	1	1	1	1	1	1	1
SMR8	6	3412	1	1	1	1	1	1	1	1
SMR9	6	1243	1	1	1	1	1	1	1	1
SMR10	6	3142	1	2	1	2	2	1	2	1
SMR11	6	2413	2	2	2	2	1	1	1	1
SMR12	6	1234	2	2	2	2	1	2	2	1
SMR13	7	4213	1	1	1	1	1	1	1	1
SMR14	7	1342	1	1	1	1	1	1	1	1
SMR15	7	4213	1	1	1	1	1	1	1	1
SMR16	7	2314	1	1	1	1	1	1	1	1
SMR17	7	4[illegible]	1	1	1	1	1	1	1	1
SMR18	7	1423	1	1	1	1	1	1	1	1
SMR19	7	2134	1	1	1	1	1	1	1	1
SMR20	7	2431	2	1	1	1	2	1	2	1
SMR21	7	1324	?	2	2	2	1	1	1	2
SMR22	7	2314	1	1	1	1	2	2	2	2
SMR23	7	2134	2	2	2	2	2	2	2	2
SMR24	7	3124	2	2	2	2	2	2	2	2
SMR25	8	3241	1	1	1	1	1	1	1	1
SMR26	8	3124	1	1	1	1	1	1	1	1
SMR27	8	2143	1	1	1	1	1	1	1	1
SMR28	8	1432	1	1	1	1	1	1	1	1
SMR29	8	2143	2	1	1	1	1	1	1	1
SMR30	8	1243	1	1	2	1	1	1	1	1
SMR31	8	4213	4	4?	1	1	2	1	1	1
SMR32	8	1342	2	1	1	2	1	2	2	2
SMR33	8	1324	2	1	2	2	1	2	2	2
SMR34	8	3421	2	2	2	2	2	2	2	?
SMR35	8	2431	2	2	2	2	2	2	2	2
SMR36	8	2143	2	2	2	2	2	2	2	2

(CONTINUED OVERLEAF)

INTERPRETATIVE CATEGORISATIONS OF MODIFIED ROULETTE EXPERIMENT RESULTS (CONT'D).

SUBJECT	AGE (YEARS)	ORDER OF PRESENTATION OF PROBLEMS	PROBLEM MR1		PROBLEM MR2		PROBLEM MR3		PROBLEM MR4	
			GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SMR37	9	3412	1	1	1	1	1	1	1	1
SMR38	9	4312	1	1	1	1	1	1	1	1
SMR39	9	2341	1	1	1	1	1	1	1	1
SMR40	9	3214	1	1	1	2	1	2	1	2
SMR41	9	4123	?	2	2	1	1	2	1	1
SMR42	9	2431	2	2	1	1	1	2	1	2
SMR43	9	1423	1	1	2	2	2	2	2	1
SMR44	9	2413	2	1	2	2	2	2	2	2
SMR45	9	3241	2	2	2	2	2	2	2	2
SMR46	9	3214	2	2	2	2	2	2	2	2
SMR47	9	1432	2	2	2	2	2	2	2	2
SMR48	9	4312	3	3	2	2	2	1	2	2
SMR49	10	1432	1	1	1	1	1	1	1	1
SMR50	10	1324	1	1	1	1	1	1	1	1
SMR51	10	2134	1	1	1	1	1	1	1	1
SMR52	10	3142	1	1	2	1	1	2	1	2
SMR53	10	1423	2	2	2	2	2	2	2	2
SMR54	10	2314	2	2	2	2	2	2	2	2
SMR55	10	2413	2	2	2	2	2	2	2	2
SMR56	10	4231	2	2	2	2	2	2	2	2
SMR57	10	2413	2	2	2	2	2	2	2	2
SMR58	10	4312	2	2	2	2	2	2	2	2
SMR59	10	3412	3	3	2	2	2	2	2	1
SMR60	10	4312	3	3	2	2	2	2	2	2

TABLE OF THE FREQUENCY OF THE DIFFERENT TYPES OF RESPONSE AT DIFFERENT AGES (MODIFIED ROULETTE EXPERIMENT).

AGE GROUP	NUMBER OF TYPE 1 RESPONSES	NUMBER OF TYPE 2 RESPONSES	NUMBER OF TYPE 3 RESPONSES	NUMBER OF TYPE 4 RESPONSES	NUMBER OF UNCLASSIFIABLE RESPONSES
6	82	14	0	0	0
7	68	27	0	0	1
8	56	37	0	2	1
9	42	51	2	0	1
10	30	62	4	0	0

APPENDIX. 11

Cluster analysis of results of the modified roulette experiment.

Contents:

TABLE OF INITIAL CATEGORISATIONS OF SUBJECTS' RESPONSES TO THE MODIFIED ROULETTE EXPERIMENT ITEMS.

SUBJECT	AGE (YEARS)	PROBLEM MR1		PROBLEM MR2		PROBLEM MR3		PROBLEM MR4	
		GREEN	RED	GREEN	RED	GREEN	RED	GREEN	RED
SMR1	6	FPRS	FQACB	FUB	FPRS	PPR	PPR	PPR	FPR
SMR2	6	FUS	FUB	FUS	FUB	FU	FU	FU	PU
SMR3	6	FUS	FPRS	FPRB	FSRS	FSR	PSR	PSR	PSR
SMR4	6	FSRB	FSRS	FSRS	FSRS	PSR	FSR	PSR	PSR
SMR5	6	FSRB	FSRS	FSRB	FSRB	PSR	PSR	PSR	PSR
SMR6	6	FRFS	FRFB	FRFS	FRFB	FRF	FRF	FRF	FRF
SMR7	6	FUB	FUS	FUS	FUB	FU	FU	PU	PU
SMR8	6	FSRS	FSRB	FSRB	FSRS	PSR	PSR	FSR	FSR
SMR9	6	FPRS	FQACTB	FQACTS	FQACTS	FPR	PPR	FQACT	PPR
SMR10	6	FQTCL	FQTC	FQTCL	FQTM	FQTECX	FQTL	FQTC	FQTCL
SMR11	6	FQTE	FQTE	FQTC	FQTECX	PSR	FSR	PRF	FSR
SMR12	6	FQTECX	FQTM	FQTM	FQTC	PSR	PQTC	FQTC	FQTCL
SMR13	7	FSRB	FSRS	FSRS	FSRB	PSR	PSR	FSR	FSR
SMR14	7	FPRB	FPRS	FPRB	FPRS	FPR	FPR	PPR	PQAPE
SMR15	7	FUS	FUB	FUB	FUS	PU	PU	PU	PU
SMR16	7	FUS	FUS	FUS	FUS	PU	PU	FU	PU
SMR17	7	FUB	FUB	FUB	FPRB	FU	PU	FU	FU
SMR18	7	FSRS	FUS	FSRS	FSRB	FSR	FSR	PSR	FSR
SMR19	7	FSRS	FSRB	FSRB	FSRB	PSR	FSR	PSR	PSR
SMR20	7	FQTC	FUB	FSRB	FSRS	FQTE	FQACT	FQTM	FSR
SMR21	7	FQAES	FQTE	FQTE	FQTC	FQAC	FU	PQAM	PQTE
SMR22	7	FQACB	FQACS	FQAPE	FQACTS	PQTENTC	PQTC	PQNTPE	PQTE
SMR23	7	FQTE	FQTE	FQTMR	FQTE	FQTE	PQTE	FQTE	PQTE
SMR24	7	FQTC	FQTEMX	FQTC	FQTC	FQTECX	PQTC	FQTC	PQTC
SMR25	8	FUS	FQACB	FQACB	FUS	PSR	FU	PQAM	PRF
SMR26	8	FUB	FUS	FSRB	FUB	FSR	PU	PU	PU
SMR27	8	FUB	FUS	FUS	FUB	PU	PU	FU	FU
SMR28	8	FSRB	FSRS	FSRB	FPRS	PSR	PSR	FPR	FU
SMR29	8	FQTFDS	FSRB	FUB	FRFB	PU	PQAPE	PSR	PQAPL
SMR30	8	FUS	FUB	FQTM	FQACTB	PSR	PSR	FSR	FSR
SMR31	8	PQRF	PQRF	FPRB	FUS	FQTPE	FU	FPR	PSR
SMR32	8	FQTC	FQTCL	FQTCL	FQTC	PSR	PQTC	FQTC	PQTC
SMR33	8	FQTC	FSRB	FQTC	FQTC	FSR	PQTC	FQTC	PQTC
SMR34	8	FQTE	FQTE	FQTMEX	FQTW	FQTE	PQTCE	FQTEM	FRF
SMR35	8	FQTPW	FQTPE	FQTPW	FQTPW	FQTW	PQTS	FQTS	PQTS
SMR36	8	FQTC	FQTC	FQTE	FQTC	PQTM	PQTEC	FQTXCXE	PQTS

(CONTINUED OVERLEAF)

TABLE OF INITIAL CATEGORISATIONS OF MODIFIED ROULETTE EXPERIMENT RESULTS (CONT'D).

SUBJECT	AGE (YEARS)	PROBLEM MR1 GREEN	PROBLEM MR1 RED	PROBLEM MR2 GREEN	PROBLEM MR2 RED	PROBLEM MR3 GREEN	PROBLEM MR3 RED	PROBLEM MR4 GREEN	PROBLEM MR4 RED
SMR37	9	FQAFB	FSRS	FQACS	FRFB	PPR	PPR	FU	FU
SMR38	9	FSRB	FUS	FSRS	FSRB	PSR	PSR	FSR	FSR
SMR39	9	FSRB	FSRS	FSRB	FSRS	PQAE	PSR	PQAM	PSR
SMR40	9	FUS	FUS	FPRB	FQTE	PU	PQTF	PPR	PQTE
SMR41	9	FQAEB	FQTE	FQTE	FUS	PQAM	PQTE	PQM	PSR
SMR42	9	FQTE	FQTE	FQACB	FPRB	PQAC	PQTCE	PQAMS	PQTM
SMR43	9	FPRS	FSRS	FQTM	FQTE	FQTE	PQTM	FQTS	FU
SMR44	9	FQTE	FSRS	FQTE	FQTE	FQTDE	PQTEC	FQTMR	PQTSR
SMR45	9	FQTE	FQTE	FQTE	FQTE	PQNTL	PQTE	PQNTE	PQTE
SMR46	9	FQTE	FQTE	FQTE	FQTE	PQNTF	PQTE	FQTE	PQTE
SMR47	9	FQTDE	FQTE	FQTDE	FQTE	PQNTE	PQTMR	PQNTL	PQTMR
SMR48	9	PQRC	PQRC	FQTC	FQTC	FQTECX	FRF	FQTC	PQTC
SMR49	10	FSRB	FUS	FUB	FUB	PQAE	FU	FU	FSR
SMR50	10	FSRS	FSRB	FSRB	FSRS	FSR	PSR	PSR	PSR
SMR51	10	FQACS	FQACTB	FPRB	FQACS	PQAC	PQAC	FQAM	FRF
SMR52	10	FQACS	FSRB	FQNTC	FSRB	FQACT	PQTC	PQAM	PQTC
SMR53	10	FQTDE	FQTDE	FQTSE	FQTSE	PQTF	PQTSE	FQTSR	PQTMR
SMR54	10	FQTECX	FQTC	FQTC	FQTECX	FQTE	PQTE	FQTS	PQTC
SMR55	10	FQTEL	FQTM	FQTE	FQTECX	FQTE	PQTE	FQTS	PQTCE
SMR56	10	PQTF	PQTF	FQTPW	FQTPE	FQTEPW	PQTEPW	FQTE	PQTC
SMR57	10	FQTM	FQTW	FQTWMX	FQTM	PQNTC	PQTM	FQTMW	PQTM
SMR58	10	PQTW	PQTW	FQACS	FQTWCX	PQNTC	PQTM	FQTM	PQTMPE
SMR59	10	PQRF	PQRF	FQTM	FQTM	PQNTC	PQTM	FQTMXE	FPR
SMR60	10	PQRF	PQRF	FQNTC	FQTC	PQNTC	PQTC	FQTC	PQTC

CATEGORY NUMBERING SCHEME FOR ANALYSIS OF RESULTS OF THE MODIFIED ROULETTE EXPERIMENT.

PROBLEM MR1				PROBLEM MR2				PROBLEM MR3				PROBLEM MR4			
GREEN		RED		GREEN		RED		GREEN		RED		GREEN		RED	
CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER	CATEGORY	NUMBER
FUB	1	FUB	26	FUB	47	FUB	68	FU	87	FU	112	FU	133	FU	154
FUS	2	FUS	27	FUS	48	FUS	69	PU	88	PU	113	PU	134	PU	160
FSRB	3	FSRB	28	FSRB	49	FSRB	70	FSR	89	FSR	114	FSR	135	FSR	161
FSRS	4	FSRS	29	FSRS	50	FSRS	71	PSR	90	PSR	115	PSR	136	PSR	162
FPRB	5	FPRS	30	FPRB	51	FPRB	72	FPR	91	FPR	116	FPR	137	FPR	163
FPRS	6	FRFB	31	FRFS	52	FPRS	73	PPR	92	PPR	117	PPR	138	PPR	164
FRFS	7	FQACB	32	FQACB	53	FRFB	74	FRF	93	FRF	118	FRF	139	FRF	165
FQACS	8	FQACS	33	FQACS	54	FQACS	75	FQACT	94	FQACT	119	PRF	140	PRF	166
FQAFB	9	FQACTB	34	FQAPE	55	FQACTB	76	FQAC	95	PQAC	120	FQACT	141	PQAPL	167
FQTCL	10	FQTCL	35	FQACTS	56	FQACTS	77	PQAM	96	PQAPE	121	FQAM	142	PQAPE	168
FQTM	11	FQTM	36	FQTCL	57	FQTM	78	PQAC	97	FQTL	122	PQM	143	FQTCL	169
FQTC	12	FQTC	37	FQTM	58	FQTC	79	PQAE	98	PQTM	123	PQAM	144	PQTM	170
FQTE	13	FQTE	38	FQTC	59	FQTE	80	FQTE	99	PQTC	124	FQAMS	145	PQTC	171
FQTPW	14	FQTDE	39	FQTE	60	FQTPE	81	FQTDE	100	PQTE	125	FQTM	146	PQTE	172
FQTDE	15	FQTPE	40	FQTMR	61	FQTSE	82	FQTPE	101	PQTS	126	FQTC	147	PQTS	173
FQTEL	16	FQTW	41	FQTDE	62	FQTPW	83	FQTW	102	PQTMR	127	FQTE	148	PQTMR	174
FQTFDS	17	FQTEMX	42	FQTSE	63	FQTW	84	FQTECX	103	PQTSE	128	FQTMR	149	PQTSR	175
FQTECX	18	PQTW	43	FQTPW	64	FQTECX	85	FQTEPW	104	PQTF	129	FQTSR	150	PQTMR	176
PQTW	19	PQTF	44	FQNIC	65	FQTWCX	86	PQNTL	105	PQTEPW	130	FQTS	151	PQTCE	177
PQTF	20	PQRC	45	FQTMEX	66			PQTM	106	PQTCE	131	FQTMW	152		
FQACB	21	PQRF	46	FQTWMX	67			PQTF	107	PQTEC	132	FQTEM	153		
FQAES	22							PQNTC	108			FQTMXE	154		
FQAEB	23							PQNTE	109			FQTXCXE	155		
PQRC	24							PQNTF	110			PQNTL	156		
PQRF	25							PQTENC	111			PQNTE	157		
												PQNTRE	158		

DATA FOR CLUSTER ANALYSIS OF RESULTS OF THE MODIFIED ROULETTE EXPERIMENT.

SUBJECT	CATEGORIES						
SMR1	6	32	47	73	92	117	138	163
SMR2	2	26	48	68	87	112	133	160
SMR3	2	30	51	71	89	115	136	162
SMR4	3	29	50	71	90	114	136	162
SMR5	3	29	49	70	90	115	136	162
SMR6	7	31	52	74	93	118	139	165
SMR7	1	27	48	68	87	112	134	160
SMR8	4	28	49	71	90	115	135	161
SMR9	6	34	56	77	91	117	141	164
SMR10	10	37	57	78	103	122	147	169
SMR11	13	38	59	85	90	114	140	161
SMR12	18	36	58	79	90	124	147	169
SMR13	3	29	50	70	90	115	135	161
SMR14	5	30	51	73	91	116	138	168
SMR15	2	26	47	69	88	113	134	160
SMR16	2	27	48	69	88	113	133	160
SMR17	1	26	47	72	87	113	133	159
SMR18	4	27	50	70	89	114	136	161
SMR19	4	28	49	70	90	114	136	162
SMR20	12	26	49	71	99	119	146	161
SMR21	22	38	60	79	95	112	144	172
SMR22	21	33	55	77	111	124	158	172
SMR23	13	38	61	80	99	125	148	172
SMR24	12	42	59	79	103	124	147	171
SMR25	2	32	53	69	90	112	144	166
SMR26	1	27	49	68	89	113	134	160
SMR27	1	27	48	68	88	113	133	159
SMR28	3	29	49	73	90	115	137	159
SMR29	17	28	47	74	88	121	136	167
SMR30	2	26	58	76	90	115	135	161
SMR31	25	46	51	69	101	112	137	162
SMR32	12	35	57	79	90	124	147	171
SMR33	12	28	59	79	89	124	147	171
SMR34	13	38	66	84	99	131	153	165
SMR35	14	40	64	83	102	126	151	173
SMR36	12	37	60	79	106	132	155	173

(CONTINUED OVERLEAF)

DATA FOR CLUSTER ANALYSIS OF MODIFIED ROULETTE EXPERIMENT RESULTS (CONT'D).

SUBJECT	CATEGORIES							
SMR37	9	29	54	74	92	117	133	159
SMR38	3	27	50	70	90	115	135	161
SMR39	3	29	49	71	98	115	144	162
SMR40	2	27	51	80	88	129	138	172
SMR41	23	38	60	69	96	125	143	162
SMR42	13	38	53	72	97	131	145	170
SMR43	6	29	58	80	99	123	151	159
SMR44	13	29	60	80	100	132	149	175
SMR45	13	38	60	80	105	125	157	172
SMR46	13	38	60	80	110	125	148	172
SMR47	15	38	62	80	109	127	156	174
SMR48	24	45	59	79	103	118	147	171
SMR49	3	27	47	68	98	112	133	161
SMR50	4	28	49	71	89	115	136	162
SMR51	8	34	51	75	97	120	142	165
SMR52	8	28	65	70	94	124	144	171
SMR53	15	39	63	82	107	128	150	174
SMR54	18	37	59	85	99	125	151	171
SMR55	16	36	60	85	99	125	151	177
SMR56	20	44	64	81	104	130	148	171
SMR57	11	41	67	78	108	123	152	170
SMR58	19	43	54	86	108	123	146	176
SMR59	25	46	58	78	108	123	154	163
SMR60	25	46	65	79	108	124	147	171

First-order groups of subjects generated by the cluster analysis of the results of the modified [illegible] experiment.

The subjects are listed [illegible] corresponding [illegible] group.

Group 1. SUR21, 48, 5, [illegible]

Group 2: SUR[illegible], [illegible], 32, 60, [illegible], 10, 52, 21.

Group 3: SUR22, [illegible], 7.

Group 4: SUR16, 37, 2, 7, [illegible], 35, 17, [illegible].

Group 5: SUR5[illegible], 55, [illegible], 45, [illegible].

Group 6: SUR19, [illegible], 5, [illegible], 10, 3, [illegible].

Group 7: SUR28, 49, [illegible], 5, [illegible], 26, [illegible].

Group 8: SUR[illegible], [illegible], 25, 27, [illegible], 7, 26, [illegible].

Group 9: SUR59, [illegible], 31.

Group 10. SUR57, 59, 52, 60.

Group 11: SUR[illegible], 56, 8, 19, 3, 1[illegible], [illegible].

Group 12: SUR[illegible], [illegible], 55, [illegible].

Group 13: SUR[illegible], 54, 52, [illegible], 6, [illegible].

Group 14: SUR11, [illegible], 18, 5, [illegible].

Group 15: SUR2[illegible], [illegible], 5, [illegible].

Group 16: SUR1[illegible], 1[illegible], [illegible].

Group 17: SUR[illegible].

Group 18: SUR[illegible], [illegible], 5, [illegible].

Group 19: SUR[illegible], 30, 15, [illegible], 27, [illegible].

Group 20: SUR[illegible], [illegible], 31, [illegible].

Group 21: SUR[illegible], 30, 5, 50, [illegible].

Group 22: SUR40, [illegible]6, [illegible], 45, [illegible], 7.

Group 23: SUR[illegible], [illegible]7, 5, [illegible].

Group 24: SUR[illegible], [illegible], 15, [illegible], 7, [illegible].

Group 25: SUR[illegible], [illegible], 3, [illegible].

Group 26: SUR[illegible], [illegible], 7, [illegible].

Group 27: SUR[illegible], 6, [illegible].

Group 28: SUR47, 37, [illegible].

Group 29: SUR[illegible].

Group 30. SUR21, 36, [illegible].

Group 31: SUR[illegible], 27, [illegible].

Group 32: SUR[illegible], [illegible], 57, [illegible].

Group 33: SUR[illegible], [illegible]7, [illegible].

Group 34: SUR[illegible].

Group 35: [illegible].

Clusters generated by the analysis of the results of the modified roulette experiment.

The groups are listed in the order generated by the analysis program, roughly corresponding to decreasing order of [illegible] to the cluster.

Cluster 1:

Groups of cluster 1: group 3, group [illegible], group [illegible], group 1[illegible], group 5, group 15, group [illegible].

Subjects of cluster 1: SMR11, [illegible]1, [illegible], [illegible], [illegible], [illegible]1, 4[illegible], [illegible], 44, [illegible]5, 46, 47.

Subjects not of cluster 1: SMR[illegible], [illegible], [illegible], 5, [illegible], 7, 8, 9, 10, 1[illegible], [illegible], 15, 17, 18, [illegible], 27, 29, [illegible], [illegible], 4[illegible], [illegible], 50, 51, 52, 5[illegible], 57, 58, 59, [illegible].

Cluster 2:

Groups of cluster 2: group 4, group 8, group 31, group 26, group 19, group 24, group 2[illegible], group 7, group 1[illegible], group 25, group 6, group 11, group 1, group 2[illegible], group 16, group 29, group 14, group 17, group 2[illegible], group 15, group 20.

Subjects of cluster 2: SMR18, 25, 26, 2[illegible], [illegible], [illegible], 4[illegible].

Subjects not of cluster 2: SMR[illegible], 1[illegible], [illegible], [illegible], 35, [illegible], 47, 53, 56, 57, 58.

Cluster 3:

Groups of cluster 3: group 9, group 10, group 27, group 2, group 17, group 14, group 21, group 1, group 11.

Subjects of cluster 3: SMR12, 3[illegible], 33, 52.

Subjects not of cluster 3: SMR[illegible], 6, 9, 14, [illegible], 3[illegible], 35, 42, 47, 51, 53, 55,

Cluster 4:

Groups of cluster 4: group 10, group 12, group 9, group 27, group 2, group 17, group 14, group 21, group 1.

Subjects of cluster 4: SMR12.

Subjects not of cluster 4: SMR[illegible], 6, 7, 9, 14, 15, 16, 17, 27, 34, 35, 40, 42, 47, 5[illegible], 53.

Cluster 5:

Groups of cluster 5: group [illegible], group 3[illegible], group [illegible], group 28, group 21, group 23.

Subjects of cluster 5: SMR1, 2[illegible], 17.

Subjects not of cluster 5: SMR[illegible], 21, [illegible], 2[illegible], 32, [illegible], 3[illegible], 36, 4[illegible], 47, 48, 5[illegible], 56.

Clusters generated by the analysis of the results of the modified roulette experiment. (Contd.)

Cluster 6:

Groups of cluster 6: group 24, group 3[illegible], group 2[illegible], group 4, group 8, group 31, group 19, group 26, group 7, group 18, group 22, group 2[illegible], group 29, group 1, group 21, group 16, group 15.

Subjects of cluster 6: SMR 2[illegible], 49.

Subjects not of cluster 6: SMR6, 10, 22, 24, 34, 35, 36, 42, 47, 48, 53, 56, 57.

Cluster 7:

Groups of cluster 7: group 12, group 32, group 13, group 5, group 30.

Subjects of cluster 7: SMR 23, 43, 44, 54, 55.

Subjects not of cluster 7: SMR 2, 3, 5, 6, 7, 8, 9, 13, 14, 15, 16, 17, 18, 26, 27, [illegible], 38, 46, 49, 50, 51, 52, 53, 56.

Cluster 8:

Groups of cluster 8: group 15, group 32, group 25, group 1, group 21, group 14, group 17, group 22, group 31.

Subjects of cluster 8: SMR20, 28, 30, 31, 37.

Subjects not of cluster 8: SMR 6, 9, 14, [illegible], 34, 35, 36, 42, 47, 51, 53, 56.

Cluster 9:

Groups of cluster 9: group 24, group 34, group 3[illegible], group 4, group 8, group 22, group 19, group 26, group 18, group 20, group 29, group 2[illegible], group 32.

Subjects of cluster 9: SMR28.

Subjects not of cluster 9: SMR6, 10, 22, 2[illegible], 34, 35, 36, 42, 47, 48, 53, 56, 57.

Cluster 10:

Groups of cluster 10: group 2, group 3[illegible], group 2[illegible], group 10, group 9, group 17, group 32.

Subjects of cluster 10: SMR12.

Subjects not of cluster 10: SMR2, 6, 7, 9, 14, 15, 16, 17, 27, [illegible], 51, 5[illegible].

Cluster 11:

Groups of cluster 11: group 2, group 16, group 1, group 21, group 14, group 17, group 25, group 6, group 2[illegible], group 7, group 13, group 2[illegible], group 11.

Subjects of cluster 11: SMR 5, 8, 11, 12, 1[illegible], 19, [illegible], 2[illegible], 3[illegible], [illegible], 39, 5[illegible], 52.

Subjects not of cluster 11: SMR6, 9, 34, 35, [illegible], 47, 51, 54, 57, 58.

Clusters generated by the analysis of the results of the modified roulette experiment. (Contd.)

Cluster 12:

Groups of cluster 12: group 18, group 34, group 35, group 29.

Subjects of cluster 12: SMR40, 43.

Subjects not of cluster 12: SMR2, 6, 7, 10, 17, 24, 27, 35, 48, 56, 57, 58, 59, 60.

Cluster 13:

Groups of cluster 13: group 5, group 27, group 1[illegible], group 30, group 2, group 22.

Subjects of cluster 13: SMR1[illegible], 54.

Subjects not of cluster 13: SMR2, 3, 6, 7, 9, 14, 15, 16, 17, 18, 26, 27, 29, 49, 51, 53.

Cluster 14:

Groups of cluster 14: group 9, group 16, group 17, group 14, group 21, group 1, group 25, group 6, group 29, group 15, group 11, group 2.

Subjects of cluster 14: SMR12, 25, 32, 50, 52.

Subjects not of cluster 14: SMR1, 6, 9, 24, 35, 48, 47, 51, 5[illegible].

Cluster 15:

Groups of cluster 15: group 10, group 32, group 3, group 23.

Subjects of cluster 15: SMR 22, 43, 44, 55.

Subjects not of cluster 51: SMR2, 6, 7, 10, 15, 16, 24, 33, 35, 48, 51, 52, 56, 57, 60.

Cluster 16:

Groups of cluster 16: group 16, group 27, group 8, group 17, group 14, group 21, group 1, group 25, group 11, group 9.

Subjects of cluster 16: SMR12, 30, 33, 52.

Subjects not of cluster 16: SMR 1, 6, 9, 14, 23, 24, 25, 48, 47, 51, 52, 55.

Cluster 17:

Groups of cluster 17: group 5, group 2[illegible], group 3, group 18, group 30.

Subjects of cluster 17: SMR11, 20, 42, 43, 44, 45, 46, 55.

Subjects not of cluster 17: SMR2, 6, 7, 9, 10, 14, 15, 16, 25, [illegible], 51, 52, 56, 57, [illegible], 59, 60.

APPENDIX I

A Technique for Classifying Qualitative Data.

by [illegible] A. [illegible] (1973).
University of Bristol,
School of Education.

A TECHNIQUE FOR CLASSIFYING QUALITATIVE DATA

This technique was developed to serve a need for an inductive procedure that would enable classes of qualitative categories to be formed and identified and which would permit such classes to make no greater assumptions about the nature of the data than that they were binary events whose frequency of occurrence and co-occurrence were subject to the laws of probability. While relative propinquity of categories in such classes might be a useful way of conceiving their multiple dependencies, there would be no necessity to invoke spatial models to express strength of coherence or the degree to which any category was bonded to the most strongly cohering category. It should be emphasised that the aim was to create a classificatory tool and to test the plausibility of the classes so formed. The resulting classes, while aiming to represent a parsimonious sorting of data would display alternative classes and would permit rational hypotheses to be fitted for subsequent experimental verification.

The development of this technique is discussed, beginning with the problem of determining the conditions for positing multiple contingencies

I. The two category and three category problem:-

1. Given three categories a, b and c where n_a ; n_b ; n_c ; are the respective frequencies of these categories in a set of N elements

a) The expected frequencies of two category possession are:-

$$n'_{ab} = \frac{n_a n_b}{N} \; ; \quad n'_{ac} = \frac{n_a n_c}{N} \; ; \quad n'_{bc} = \frac{n_b n_c}{N}$$

2. Given that n_{ab}, n_{ac} and n_{bc} are known, the probability of choosing from those elements that are b an element which is also a is $\frac{n_{ab}}{n_b}$, and the probability of choosing from those elements that are b one that is also c is $\frac{n_{bc}}{n_b}$, and the probability of choosing an element that is a and c from those that are b is $\frac{n_{ab} \times n_{bc}}{n_b^2}$; the expected value of n_{abc} being $\frac{n_{ab} \times n_{bc}}{n_b}$.

$$n'_{abc} = \frac{n_{ab} \times n_{bc}}{n_b} \quad \text{.............................(1)}$$

Similarly: expected values of n_{abc} can be derived from

$$\frac{n_{ab} \times n_{ac}}{n_a} \quad \text{and} \quad \frac{n_{ad} \times n_{bc}}{n_c}.$$

3. If $n_{ab} = \frac{n_a n_b}{N}$ and $n_{bc} = \frac{n_c n_c}{N}$

then from (1) $\dfrac{\frac{n_a n_b}{N} \times \frac{n_b n_c}{N}}{n_b} = n'_{abc} = \dfrac{n_a n_b n_c}{N^2}$.

4. It follows that if $n_{ab} > \frac{n_a n_b}{N}$ and $n_{bc} > \frac{n_b n_c}{N}$,

then $\dfrac{n_{ab} \times n_{bc}}{n_b} > \dfrac{n_a n_b n_c}{N^2}$,

and it has been shown that if amongst three given categories the obtained frequency of joint belonging to any two pairs is greater than expected by random assignment, the frequency of joint belonging to all three categories will be estimated to exceed the frequency expected from random assignment to all three categories independently.

It follows that if all three pairs of frequencies of joint belonging exceed expected values, the hypothesis that

$n_{abc} > \dfrac{n_a n_b n_c}{N^2}$ is strengthened.

II. Higher multiple category problems:-

1. Any number of categories greater than three can be dealt with by extrapolation of this argument since they can be regarded as reducible to all possible paired relationships between categories. It will suffice to demonstrate this in a four category problem.

2. a, b, c, d are four categories for which it has been found that

$$n_{ab} > \frac{n_a n_b}{N} \; ; \quad n_{ac} > \frac{n_a n_c}{N} \; ; \quad n_{ad} > \frac{n_a n_d}{N} \; ; \quad n_{bc} > \frac{n_b n_c}{N}$$

.... to all possible pairs.

If they were merely equal then the probability of an element which is b being a, c and d is

$\frac{n_{ab}}{b} \times \frac{n_{bc}}{b} \times \frac{n_{bd}}{b}$ and the expected frequency in n_b

elements would be $\frac{n_{ab} \times n_{bc} \times n_{bd}}{n_b^2}$, which after substitution

becomes $\frac{n_a n_b n_c n_d}{N^3}$.

This is the frequency that would be expected from independent random assignment to a, b, c and d,

but since $n_{ab} > \frac{n_a n_b}{N}$; $n_{bc} > \frac{n_b n_c}{N}$; $n_{bd} > \frac{n_b n_d}{N}$, then

$$\frac{n_{ab} \times n_{bc} \times n_{bd}}{n_b^2} > \frac{n_a n_b n_c n_d}{N^2} .$$

When this is demonstrable for all possible triadic relationships between categories, then the hypothesis that

$n_{abcd} > \frac{n_a n_b n_c n_d}{N^2}$ is supported.

III. <u>The problem of identifying those categories amongst which mutual belonging of elements exceeds that expected by random assignment to categories.</u>

A first step in tackling this problem would be to determine for which categories obtained frequencies of belonging to each pair of every possible pair exceeded expectation. If our concern is to make the most economical description of co-variation between categories or to assert that the number of categories in a group has been increased to the limit required by $f_o > f_e$ then a procedure must be devised for identifying that maximal set of categories in which the observed frequency of joint belonging to pairs of categories is greater than the expected frequency for all possible pairs.

An iterative procedure suggests itself even though it may not be the most economical strategy. If we begin by selecting that pair of categories for which $f_o - f_e$ is greatest we increase the likelihood that another category can be found for which joint belonging to each member of the pair is also in each case greater than expected. The choice of this third category would on the same principle be made by selecting that for which the sum of $f_o - f_e$ for joint belonging to each member of the first pair is greatest. This process can be carried on until a point is reached where no further category can be added which satisfies the $f_o > f_e$ criterion. The order in which categories are added to a group, being determined by the order of the sums of differences, contributes to a maximising of the size of the group formed. Further groups can be formed in the same way until every category has been considered for its contribution to a group and has been assigned to a group or identified as having no co-variation with any other category which is greater than chance.

IV. The problem of standardising frequencies and differences between obtained and expected frequencies.

1. Such a process as that described in III could lead to bias in the groups formed as a result of the variation in the number of elements in each category. It is desirable, therefore, to find some way in which the raw frequencies can be standardised. In a symmetrical matrix of categories in which the elements of the matrix are frequencies of joint belonging to categories i and j the frequency may be written n_{ij}. It will be assumed for the present case that the categories are not mutually exclusive and that a total of

N elements are assigned independently to categories i and j. The raw frequency n_{ij} may be expressed as the proportion,

$$p_{ij} = \frac{n_{ij}}{N}$$

and the proportion itself standardised by expressing it as a ratio of the standard deviation of the proportion *,

$$\sigma\, p_{ij} = \sqrt{\frac{n_{i}}{N} \times \frac{n_{j}}{N}}$$

where n_{i} and n_{j} are respectively the frequencies of belonging of elements to i and j the obtained standardised relative frequency then becomes

$$\frac{p_{ij}}{\sigma\, p_{ij}} = \frac{n_{ij}}{\sqrt{n_i n_j}}$$

2. The problem of standardising differences between obtained and expected frequencies can be regarded in the following way, where

n_{ij} = the observed raw frequency

$\frac{n_i n_j}{N}$ = the expected raw frequency.

The ratio of obtained to expected frequency is

$$\frac{n_{ij}}{n_i n_j} \times \frac{N}{1} = \frac{N}{\sqrt{n_i n_j}} \times \frac{n_{ij}}{\sqrt{n_i n_j}}$$

If the obtained and expected frequencies were equal, then

$$\frac{N}{\sqrt{n_i n_j}} \times \frac{n_{ij}}{\sqrt{n_i n_j}} = 1 \quad \text{and} \quad \frac{n_{ij}}{\sqrt{n_i n_j}} = \sqrt{\frac{n_i n_j}{N}},$$

whence $$\frac{n_{ij}}{\sqrt{n_i n_j}} - \sqrt{\frac{n_i n_j}{N}} = 0.$$

* See Burt, C., "Statistical Problems in the Evaluation of Army Tests", Psychometrica, IX, 1944, p. 225.

Now the first term of the above equation has been identified (2) as the observed, standardised relative frequency, therefore the second term must represent the expected, standardised relative frequency. And the whole expression is the standardised difference between observed and expected frequencies.

3. A further demonstration of this can be found by examining the relationship of the above equation to X^2

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e}.$$

which in the notation we have used becomes

$$X^2 = \sum \frac{\left[n_{ij} - \frac{n_i n_j}{N}\right]^2}{\frac{n_i n_j}{N}}$$

If we consider the square root of the term

$$\frac{\left[n_{ij} - \frac{n_i n_j}{N}\right]^2}{\frac{n_i n_j}{N}},$$

it can be expressed as $\left[n_{ij} - \frac{n_i n_j}{N}\right] \times \frac{\sqrt{N}}{\sqrt{n_i n_j}}$

whence $\frac{\sqrt{N}\, n_{ij}}{\sqrt{n_i n_j}} - \frac{\sqrt{n_i n_j}}{\sqrt{N}}$, and $\sqrt{N}\left[\frac{n_{ij}}{\sqrt{n_i n_j}} - \frac{\sqrt{n_i n_j}}{N}\right]$

and $X^2 = N\, \Sigma \left[\frac{n_{ij}}{\sqrt{n_i n_j}} - \frac{\sqrt{n_i n_j}}{N}\right]^2$

the term inside the brackets in the equation above is identical with that we have described as the standardised difference between observed and expected frequencies. A matrix of these values has been shown by Burt (1953) to be the residual matrix after the abstraction of the first general factor from the matrix of standardised relative frequencies.

V. The problem of mutually exclusive categories

A matrix of frequencies of the type we have discussed might refer to categories some of which are mutually exclusive and others not. The question that arises is whether or not the proposed use of standardised relative differences is compatible with such a matrix. Since amongst any set of mutually exclusive categories, observed frequencies are zero and, by definition, no frequency of joint belonging to any pair of categories can occur, the first question is whether or not a notion of expected frequency is meaningful.

It also follows that if the value $\frac{n_i n_j}{N}$ is subtracted from zero and a negative difference obtained, this negative value will be treated as having the same meaning as a negative difference arising from non-mutually exclusive categories. These problems may be examined by considering the case of zero observed frequency of joint belonging to
a) a pair of mutually exclusive categories
b) a pair of non-mutually exclusive categories.
In the case of b) it is a premiss of the matrix that the frequency may be non-zero although it is a necessary condition of the operation to be performed on the matrix that it admit of this premiss being false.
In the case of a) this premiss is known to be false. Thus so long as the operations performed on the matrix are such that they admit and identify cases where the premiss is or may be false and do not argue from these false premisses, both mutually exclusive and non-mutually exclusive categories are admissible in the same matrix. The function of an expected frequency in the cases we have discussed is to create a difference which will lead to the identification of those cases where the premiss may be false.

Another way of regarding negative differences in both mutually exclusive and non-mutually exclusive categories is that they express the degree to which the possession of one attribute implies the absence of the other.

VI The problem of partial association with a group of categories

1. The method proposed in III for identifying multiple category groups is in itself stringent in that it requires that all possible pairs of categories satisfy the condition $f_o > f_e$, when less than this condition

might be sufficient in any significantly related set of categories. The effect of the condition might be to exclude a category for which association with the group might be statistically significant. This can be readily seen in the case where one of a set of mutually exclusive categories has been admitted to a group and by its nature no further category in that set can be admitted. Yet the relationship of the group with two or more categories of the set might be disjunctive i.e. either a or b. Certain other pairs of categories might have the same practical status although they have not been logically constructed to have that status. Thus two or more non-mutually exclusive categories may in practice, operate to exclude one another, yet possess a disjunctive relationship with the group. While in the case of logically, mutually-exclusive categories it would be a simple matter to take their status into account, in the case of empirical mutually exclusive categories difficulties would arise.

2. It seems preferable therefore to create another procedure whereby the excessive stringency of the first procedure can be corrected. In the cases we have discussed, consideration of a new category for admission to a group would reveal both positive and negative differences with constituent categories of the group. The rule that has been adopted is that if the sum of these differences is positive, there is a prima facie case for examining the relationship of that category with the group formed. The rule is somewhat arbitrary but it will be shown that it is in practice trustworthy. Thus, after the completion of any group, the sum of the differences for every category with members of the group is found.

VII. The problem of elements in the leading diagonal

The matrix of standardised relative frequencies leads to all elements in the leading diagonal being 1. After subtraction of the 'expected frequency' it becomes less than 1. The value of this 'expected frequency' $\left[\frac{n_i}{N}\right]$ is in fact equal to the probability of occurrence of that category. Thus the diagonal element is the complement of a category's probability of occurrence. The diagonal element can be used in one of two ways.

a) The substantive use in the programme as currently used

When a group has been formed, the sum of the differences for all possible pairs of the 'a' categories forming the group into which that category enters, is composed of a - 1 differences, whereas all categories not admitted to the group have sums composed of 'a' differences. Relative magnitudes of 'in - group' and 'out - group' sums would thus be non comparable and some controlling operations, like the use of the means of differences, would be necessary. An alternative procedure, and the one adopted in the programme, is to add the diagonal element for a category to its sum. The diagonal element can be considered as expressing the likelihood that any contingent event into which the category enters entails all incidences of that category. When a category with a diagonal element approaching 1 enters a group it carries with it the likelihood of high redundancy. By adding the diagonal element to the sum of differences for a category, the sum is increased by a value proportionate to the category's likely redundancy.

b) The use of diagonal elements as weights for sums of differences

The use of the diagonal elements, as discussed in a) has the effect of assigning greater importance in a group to those categories which have lower likelihood of supporting general propositions about the whole sample of objects to which categories are assigned. The justification for this lies in the intention of the classification procedure to distinguish cohering groups of categories whatever their level of generality. If, however, the intention were to identify the categories contributing the most general statement about the sample of objects, then the complement of the diagonal elements may be used to weight the sums. While such a procedure does not affect the sorting of categories into classes, it affects the relative magnitude of sums. Moreover, it requires the use of the means of weighted sums.

VIII. The status of the groups so formed

It will be evident from the earlier discussion that groups have been formed by reference to a necessary but insufficient condition. Firstly, the sorting operation utilizes only positive differences to initiate and add members to groups; secondly, no reference is made to the standard error of the differences in selecting categories for admission; thirdly, the argument for multiple inter-dependence of categories is based upon consideration of all possible pairs of categories only. The implications of these limitations are:

a. There is no guarantee that the frequency of joint belonging to all categories of a group will exceed chance expectation.

b. A group may be formed beginning with a non significant difference and grow by adding categories that also betray non significant differences.

c. The order in which categories are added and thus the power given to them to exclude other categories is based upon a difference between sums which takes no account of their standard error.

Whereas in logic we should require of a class that there be defining properties of the class such that inclusion and exclusion would be unequivocal, in psychology we frequently seek to classify in such a manner that persons who betray related but not identical properties reveal the common pattern underlying the overt personal manifestations. If our observations of persons were always grouped by the functions which gave rise to the observations, then class logic might be a relevant guide. However, insofar as the same and underlying function may give rise to different behavioural manifestations in different persons, we would by the operation of class logic be led to reject an hypothesis of the operation of that function, by failing to find significant joint belonging to all overt categories.

Given the risk that groups may be formed which are the outcome of sampling error, the options open are either to institute step by step tests of significance or to apply an overall test of significance to each group. The former would, at the very best, substantially increase the complexity of the analysis and the latter would entail the risk of rejecting groups on the basis of what is considered to be destructive of useful hypothesis formulation. An alternative procedure is cross validation in which the analysis is applied to independent samples and groups are then tested for their replication. This is the procedure currently adopted in the programme and involves running the analysis again on the groups derived from separate samples and admitting only those groups which are replicated. This has the further virtue of revealing group redundancy which is an inevitable feature of this classificatory process.

It is a common experience in employing the process on real data that the groups formed are immediately apprehended in the context of a priori hypotheses. Nevertheless, it is worth entering a warning that the technique is not intended as a substitute for the prior rational development of hypotheses. It can act as a device for clarifying the operational implications of the a priori hypotheses and it can at times suggest counter hypotheses that have not been previously entertained. Sometimes the latter are salutory in suggesting the degree to which pre-structuring of data largely determines the outcome.

The technique has been fully programmed for the Bristol University computer and contains a number of packages. Earlier difficulties in communicating the programme, arising from writing it in machine language have been overcome by its present PORTRAN form.

Sample Analysis

For illustrative purposes the technique is shown in application to data gathered by Burt (1952) and has the advantage that a complete set of significance tests can be applied to test the results of the analysis which can also be compared with those derived from Burt's factorial approach. Table I lists the raw frequencies and Table II is the matrix of standardised differences.

The data refer to the classification of individuals in terms of Hair-colour (Fair, Red, Dark); Eye-colour (Light, Mixed, Brown); Width of Head (Narrow, Wide); Height (Tall, Short). The steps in the process of forming groups are given below.

1. Identify the highest* positive value excluding diagonal elements, in Table II (T.L. .393).

2. Identify the categories which have positive values with the two categories so selected and from these select that which has the highest* sum of values F. .365).

3. From the remaining select those which have a positive value for F and repeat from 2.

4. When the stage is reached that no further categories have positive values with all those included, the formation of the group is complete. (Neither R nor N satisfy this condition and the group is composed of T, L and F.)

5. Since condition 4 is stringent, it is preferable to add a further stage so that all promising hypotheses are entertained. For all remaining categories determine the sum of values. Accept as associated with the group all categories for which the sum is positive. (Thus R being a mutually exclusive category with F would be so associated and would lead to the hypothesis that T, L, F/R was a significant multiple combination, while N would also be tested for its association with T, L and F.) The completion of this first group suggests the following combinations as fully significant.

   1) T.L.F.
   2) T.L.F/R.
   3) T.L.F.N.

* It infrequently happens that two or more values are equal, in such cases the practice is adopted of selecting that category which has the highest sum of differences with all other categories. If two still remain, the categories are accepted in the order in which they are listed.

6. The categories not so far selected for inclusion in a central group are then explored to determine the highest positive value (BS .222).

7. Repeat as from 2

    4) B.S.D.
    5) B.S.D.M.
    6) W.M.S.
    7) W.M.S.D.
    8) R.L.T.N.
    9) R/F.L.T.N.

Nine combinations of 3 categories or more are thus identified out of the 192 that are entertained and all are significant when tested against the complete table of combinations given by Burt. No other combinations attain significance.

M.A. Brimer
March 1973.

TABLE I

F	R	D	Tot.	L	M	B	Tot.	N	W	Tot.	T	S	Tot.	Trait
														HAIR
22	0	0	22	14	6	2	22	14	8	22	13	9	22	Fair
0	15	0	15	8	5	2	15	11	4	15	10	5	15	Red
0	0	63	63	11	25	27	63	44	19	63	20	43	63	Dark
22	15	63	100	33	36	31	100	69	31	100	43	57	100	Total
														EYES
14	8	11	33	33	0	0	33	27	6	33	29	4	33	Light
6	5	25	36	0	36	0	36	20	16	36	10	26	36	Mixed
2	2	27	31	0	0	31	31	22	9	31	4	27	31	Brown
22	15	63	100	33	36	31	100	69	31	100	43	57	100	Total
														HEAD
14	11	44	69	27	20	22	69	69	0	69	30	39	69	Narrow
8	4	19	31	6	16	9	31	0	31	31	13	18	31	Wide
22	15	63	100	33	36	31	100	69	31	100	43	57	100	Total
														STATURE
13	10	20	43	29	10	4	43	30	13	43	43	0	43	Tall
9	5	43	57	4	26	27	57	39	18	57	0	57	57	Short
22	15	63	100	33	36	31	100	69	31	100	43	57	100	Total

TABLE II

F	R	D	L	M	D	N	W	T	S	Trait
										HAIR
.780	-.182	-.372	.250	-.068	-.185	-.030	.045	.115	-.100	Fair
-.182	.850	-.370	.137	-.017	.123	.020	-.030	.140	-.121	Red
-.372	-.370	.370	-.215	.049	.169	.008	-.012	-.136	.118	Dark
										EYES
.250	.137	-.215	.670	-.345	-.320	.089	-.132	.393	-.341	Light
-.068	-.017	.049	-.345	.640	-.334	-.097	.145	-.139	.121	Mixed
-.185	-.123	.169	-.320	-.334	.690	.013	-.020	-.255	.222	Brown
										HEAD
-.030	.020	.008	.089	-.097	.013	.310	-.462	.006	-.005	Narrow
.045	-.030	-.012	-.132	.145	-.020	-.462	.690	-.009	.008	Wide
										STATURE
.115	.140	-.136	.393	-.139	-.255	.006	-.009	.570	-.495	Tall
-.100	-.121	.118	-.241	.121	.222	-.005	.008	-.495	.430	Short

APPENDIX J

Statistical Tests

Contents:

Page:

STAGE CATEGORISATIONS OF SUBJECTS AGED SIX TO TEN YEARS FROM THE FIRST BEADS, ROULETTE AND MODIFIED ROULETTE EXPERIMENTS, TO PROVIDE DATA FOR THE CHI-SQUARE TEST.

THE SUBJECTS ARE ASSIGNED TO THE STAGE THAT MOST OF THEIR ANSWERS ARE FROM. TIES ARE RESOLVED BY ASSIGNING THE SUBJECT TO THE LOWER OF THE TIED STAGES.

BEADS EXPERIMENT, B1-B4.			
SUBJECT	STAGE	SUBJECT	STAGE
SB13	1	SB44	2
SB14	1	SB45	3
SB15	1	SB46	3
SB16	1	SB47	3
SB17	1	SB48	3
SB18	2	SB49	1
SB19	2	SB50	2
SB20	2	SB51	2
SB21	2	SB52	2
SB22	2	SB53	2
SB23	2	SB54	2
SB24	2	SB55	3
SB25	1	SB56	3
SB26	1	SB57	3
SB27	2	SB58	3
SB28	2	SB59	3
SB29	2	SB60	3
SB30	2	SB61	2
SB31	2	SB62	2
SB32	2	SB63	2
SB33	2	SB64	2
SB34	3	SB65	3
SB35	3	SB66	3
SB36	2	SB67	3
SB37	1	SB68	3
SB38	2	SB69	3
SB39	2	SB70	3
SB40	2	SB71	2
SB41	2	SB72	3
SB42	2		
SB43	2		

ROULETTE EXPERIMENT, R2-R5.			
SUBJECT	STAGE	SUBJECT	STAGE
SR1	1	SR32	1
SR2	1	SR33	1
SR3	1	SR34	2
SR4	1	SR35	2
SR5	1	SR36	2
SR6	1	SR37	1
SR7	1	SR38	1
SR8	1	SR39	1
SR9	1	SR40	2
SR10	1	SR41	2
SR11	1	SR42	2
SR12	1	SR43	2
SR13	1	SR44	3
SR14	1	SR45	2
SR15	1	SR46	2
SR16	1	SR47	3
SR17	1	SR48	2
SR18	1	SR49	1
SR19	1	SR50	2
SR20	1	SR51	2
SR21	1	SR52	2
SR22	1	SR53	2
SR23	2	SR54	2
SR24	3	SR55	2
SR25	1	SR56	2
SR26	1	SR57	2
SR27	1	SR58	2
SR28	1	SR59	2
SR29	1	SR60	3
SR30	1		
SR31	1		

MODIFIED ROULETTE EXPERIMENT			
SUBJECT	STAGE	SUBJECT	STAGE
SMR1	1	SMR32	2
SMR2	1	SMR33	2
SMR3	1	SMR34	2
SMR4	1	SMR35	2
SMR5	1	SMR36	2
SMR6	1	SMR37	1
SMR7	1	SMR38	1
SMR8	1	SMR39	1
SMR9	1	SMR40	1
SMR10	1	SMR41	1
SMR11	1	SMR42	1
SMR12	2	SMR43	2
SMR13	1	SMR44	2
SMR14	1	SMR45	2
SMR15	1	SMR46	2
SMR16	1	SMR47	2
SMR17	1	SMR48	2
SMR18	1	SMR49	1
SMR19	1	SMR50	1
SMR20	1	SMR51	1
SMR21	2	SMR52	1
SMR22	1	SMR53	2
SMR23	2	SMR54	2
SMR24	2	SMR55	2
SMR25	1	SMR56	2
SMR26	1	SMR57	2
SMR27	1	SMR58	2
SMR28	1	SMR59	2
SMR29	1	SMR60	2
SMR30	1		
SMR31	1		

CHI-SQUARE TEST FOR TWO INDEPENDENT SAMPLES, USED TO COMPARE THE RESULTS OF THE FIRST BEADS AND ROULETTE EXPERIMENTS.

	BEADS EXPERIMENT, B1-B4.	ROULETTE EXP'T, R2-R5.	TOTAL.
STAGE 1.	9 E.22	35 E.22	44
STAGE 2.	32 E 26·5	21 E 26·5	53
STAGE 3.	19 E 11·5	4 E 11·5	23
TOTAL	60	60	120

$$\chi^2 = \frac{(9-22)^2}{22} + \frac{(35-22)^2}{22} + \frac{(32-26\cdot5)^2}{26\cdot5} + \frac{(21-26\cdot5)^2}{26\cdot5} + \frac{(19-11\cdot5)^2}{11\cdot5} + \frac{(4-11\cdot5)^2}{11\cdot5}$$

$$\therefore \chi^2 = \underline{\underline{27\cdot4}}$$

SIGNIFICANT AT 0·001 LEVEL (D.F.=2).

CHI-SQUARE TEST FOR TWO INDEPENDENT SAMPLES, USED TO COMPARE THE RESULTS OF THE FIRST BEADS AND MODIFIED ROULETTE EXPERIMENTS

	BEADS EXPERIMENT B1-B4.	MODIFIED ROULETTE EXPERIMENT	TOTAL.
STAGE 1	9 E 23	37 E 23	46
STAGE 2	32 E 27·5	23 E 27·5	55
STAGE 3	19 E 9·5	0 E 9·5	19
TOTAL.	60	60	120

$$\chi^2 = \frac{(9-23)^2}{23} + \frac{(37-23)^2}{23} + \frac{(32-27\cdot5)^2}{27\cdot5} + \frac{(23-27\cdot5)^2}{27\cdot5} + \frac{(19-9\cdot5)^2}{9\cdot5} + \frac{(0-9\cdot5)^2}{9\cdot5}$$

$$\therefore \chi^2 = \underline{\underline{37\cdot5}}.$$

SIGNIFICANT AT 0·001 LEVEL (D.F.=2)

NUMBER OF TYPE 1, 2, AND 3 ANSWERS PRODUCED BY SUBJECTS AGED SIX TO TEN YEARS IN THE MAIN EXPERIMENTS, TO PROVIDE DATA FOR MEDIAN TESTS.

FIRST BEADS EXPERIMENT, B1-B4.			
SUBJECT	NUMBER OF TYPE 1 ANSWERS	NUMBER OF TYPE 2 ANSWERS	NUMBER OF TYPE 3 ANSWERS
SB13	8	0	0
SB14	6	2	0
SB15	6	2	0
SB16	6	2	0
SB17	5	3	0
SB18	2	5	0
SB19	1	7	0
SB20	0	6	0
SB21	0	8	0
SB22	0	8	0
SB23	0	8	0
SB24	0	7	1
SB25	8	0	0
SB26	8	0	0
SB27	2	3	0
SB28	0	8	0
SB29	0	8	0
SB30	0	8	0
SB31	0	8	0
SB32	0	8	0
SB33	0	8	0
SB34	0	2	3
SB35	0	1	5
SB36	0	4	4
SB37	7	1	0
SB38	0	8	0
SB39	0	8	0
SB40	0	8	0
SB41	0	8	0
SB42	0	8	0
SB43	0	8	0

FIRST BEADS EXPERIMENT, B1-B4			
SUBJECT	NUMBER OF TYPE 1 ANSWERS	NUMBER OF TYPE 2 ANSWERS	NUMBER OF TYPE 3 ANSWERS
SB44	0	6	2
SB48	0	3	5
SB45	0	2	6
SB46	0	1	7
SB47	0	0	8
SB49	8	0	0
SB50	2	4	0
SB51	1	5	0
SB52	0	8	0
SB53	0	8	0
SB54	0	1	4
SB55	0	4	3
SB56	0	3	5
SB57	0	2	6
SB58	0	2	6
SB59	0	2	6
SB60	0	1	7
SB61	0	8	0
SB62	0	8	0
SB63	0	8	0
SB64	0	5	0
SB71	0	4	4
SB65	0	3	5
SB66	0	1	6
SB67	0	0	8
SB68	0	0	8
SB69	0	0	8
SB70	0	0	8
SB72	0	0	8
TOTAL	70	254	133

(CONTINUED OVERLEAF).

DATA FOR MEDIAN TESTS (CONT'D).

ROULETTE EXPERIMENT, R2-R5.			
SUBJECT	NUMBER OF TYPE 1 ANSWERS	NUMBER OF TYPE 2 ANSWERS	NUMBER OF TYPE 3 ANSWERS
SR1	8	0	0
SR2	8	0	0
SR3	8	0	0
SR4	8	0	0
SR5	8	0	0
SR6	8	0	0
SR7	8	0	0
SR8	8	0	0
SR9	8	0	0
SR10	8	0	0
SR11	8	0	0
SR12	5	0	3
SR13	8	0	0
SR14	8	0	0
SR15	8	0	0
SR16	8	0	0
SR17	8	0	0
SR18	8	0	0
SR19	8	0	0
SR20	8	0	0
SR21	7	1	0
SR22	3	3	0
SR23	1	5	2
SR24	0	2	6
SR25	8	0	0
SR26	8	0	0
SR27	8	0	0
SR28	8	0	0
SR29	8	0	0
SR30	8	0	0
SR31	4	4	0

ROULETTE EXPERIMENT, R2-R5.			
SUBJECT	NUMBER OF TYPE 1 ANSWERS	NUMBER OF TYPE 2 ANSWERS	NUMBER OF TYPE 3 ANSWERS
SR32	4	2	2
SR33	4	0	4
SR34	1	6	0
SR35	0	6	2
SR36	0	6	2
SR37	5	3	0
SR38	3	2	2
SR39	3	5	0
SR40	1	5	1
SR41	2	6	0
SR42	2	5	1
SR43	1	7	0
SR44	1	2	5
SR45	0	7	1
SR46	0	6	2
SR47	0	2	6
SR48	0	6	2
SR49	8	0	0
SR50	2	6	0
SR51	3	5	0
SR52	2	6	0
SR53	2	6	0
SR54	2	5	0
SR55	1	7	0
SR56	1	6	1
SR57	0	7	0
SR58	0	5	3
SR59	0	3	5
SR60	0	0	8
TOTAL	268	147	58

(CONTINUED OVERLEAF)

DATA FOR MEDIAN TESTS (CONCLUDED).

MODIFIED ROULETTE EXPERIMENT.			
SUBJECT	NUMBER OF TYPE 1 ANSWERS	NUMBER OF TYPE 2 ANSWERS	NUMBER OF TYPE 3 ANSWERS
SMR1	8	0	0
SMR2	8	0	0
SMR3	8	0	0
SMR4	8	0	0
SMR5	8	0	0
SMR6	8	0	0
SMR7	8	0	0
SMR8	8	0	0
SMR9	8	0	0
SMR10	4	4	0
SMR11	4	4	0
SMR12	2	6	0
SMR13	8	0	0
SMR14	8	0	0
SMR15	8	0	0
SMR16	8	0	0
SMR17	8	0	0
SMR18	8	0	0
SMR19	8	0	0
SMR20	5	3	0
SMR21	3	4	0
SMR22	4	4	0
SMR23	0	8	0
SMR24	0	8	0
SMR25	8	0	0
SMR26	8	0	0
SMR27	8	0	0
SMR28	8	0	0
SMR29	7	1	0
SMR30	7	1	0
SMR31	5	1	0

MODIFIED ROULETTE EXPERIMENT.			
SUBJECT	NUMBER OF TYPE 1 ANSWERS	NUMBER OF TYPE 2 ANSWERS	NUMBER OF TYPE 3 ANSWERS
SMR32	3	5	0
SMR33	2	6	0
SMR34	0	7	0
SMR35	0	8	0
SMR36	0	8	0
SMR37	8	0	0
SMR38	8	0	0
SMR39	8	0	0
SMR40	5	3	0
SMR41	4	3	0
SMR42	4	4	0
SMR43	3	5	0
SMR44	1	7	0
SMR45	0	8	0
SMR46	0	8	0
SMR47	0	8	0
SMR48	1	5	2
SMR49	8	0	0
SMR50	8	0	0
SMR51	8	0	0
SMR52	5	3	0
SMR53	0	8	0
SMR54	0	8	0
SMR55	0	8	0
SMR56	0	8	0
SMR57	0	8	0
SMR58	0	8	0
SMR59	1	5	2
SMR60	0	6	2
TOTAL	278	191	6

MEDIAN TEST: TYPE 1 ANSWERS IN THE FIRST BEADS AND ROULETTE EXPERIMENTS

$$\text{MEDIAN} = \frac{70+268}{120} = \frac{338}{120}$$

$\therefore$ SCORES OF 3 OR OVER ARE ABOVE MEDIAN.

	BEADS EXPERIMENT	ROULETTE EXPERIMENT	TOTAL
NUMBER OF SCORES ABOVE COMBINED MEDIAN	9	36	45
NUMBER OF SCORES BELOW COMBINED MEDIAN	51	24	75
TOTAL	60	60	120

$$\therefore \chi^2 = \frac{120\left[(9\times 24 - 36\times 51) - \frac{120}{2}\right]^2}{45\times 75\times 60\times 60} = \frac{(216-1836-60)^2}{101250}$$

$$\therefore \chi^2 = \underline{\underline{27.9}}.$$

THIS VALUE IS <u>SIGNIFICANT AT THE 0.001 LEVEL (D.F=1)</u>.

MEDIAN TEST: TYPE 3 ANSWERS IN THE FIRST BEADS AND ROULETTE EXPERIMENTS.

$$\text{MEDIAN} = \frac{133+58}{120} = \frac{191}{120}$$

$\therefore$ SCORES OF 2 OR OVER ARE ABOVE MEDIAN.

	BEADS EXPERIMENT	ROULETTE EXPERIMENT	TOTAL
NUMBER OF SCORES ABOVE COMBINED MEDIAN.	23	12	35
NUMBER OF SCORES BELOW COMBINED MEDIAN.	37	48	85
TOTAL	60	60	120

$$\therefore \chi^2 = \frac{120\left[(23\times 48 - 12\times 37) - \frac{120}{2}\right]^2}{35\times 85\times 60\times 60} = \frac{[1104-444-60]^2}{89250}$$

$$\therefore \underline{\underline{\chi^2 = 4.03}}$$

PROBABILITY OF OCCURRENCE UNDER $H_0 < \frac{1}{2}\,0.005$ (D.F=1, ONE-TAILED TEST).

i.e. VALUE IS <u>SIGNIFICANT AT 0.025 LEVEL.</u>

MEDIAN TEST: TYPE 2 ANSWERS IN THE ROULETTE AND MODIFIED ROULETTE EXPERIMENTS (COMBINED 6-10 YEAR AGE GROUPS)

$$\text{MEDIAN} = \frac{147+191}{120} = \frac{338}{120}$$

∴ SCORES OF 3 OR MORE ARE ABOVE MEDIAN.

	ROULETTE EXPERIMENT	MODIFIED ROULETTE EXPERIMENT	TOTAL
NUMBER OF SCORES ABOVE COMBINED MEDIAN	25	31	56
NUMBER OF SCORES BELOW COMBINED MEDIAN.	35	29	64
TOTAL	60	60	120

$$\therefore \chi^2 = \frac{120\left[(25\times 29 - 31\times 35) - \frac{120}{2}\right]^2}{64\times 56\times 60\times 60} = \frac{\left[725-1085-60\right]^2}{107520}$$

$$\therefore \underline{\chi^2 = 1\cdot 64}$$

THIS VALUE IS NOT SIGNIFICANT (ONE-TAILED TEST, D.F = 1).

MEDIAN TEST: TYPE 2 ANSWERS IN THE ROULETTE AND MODIFIED ROULETTE EXPERIMENTS (6 AND 7 YEAR OLD SUBJECTS ONLY).

$$\text{MEDIAN} = \frac{11+41}{24} = \frac{55}{24}$$ ∴ SCORES OF 3 OR MORE ARE ABOVE MEDIAN

	ROULETTE EXPERIMENT	MODIFIED ROULETTE EXP'T.	TOTAL
NUMBER OF SCORES ABOVE COMBINED MEDIAN.	2	8	10
NUMBER OF SCORES BELOW COMBINED MEDIAN.	22	16	38
TOTAL	24	24	48

SINCE NONE OF THE EXPECTED FREQUENCIES IS LESS THAN 5, AND $N_1 + N_2 > 20$, WE MAY USE THE χ^2 TEST TO TEST H_0.

$$\chi^2 = \frac{48\left[(2\times 16 - 8\times 22) - \frac{48}{2}\right]^2}{10\times 38\times 24\times 24} = \frac{(32-176-24)^2}{4560}$$

$$\therefore \underline{\chi^2 = 6\cdot 19}.$$

PROBABILITY OF OCCURRENCE UNDER $H_0 < \frac{1}{2}(0\cdot 01)$ (D.F.=1, ONE-TAILED TEST)
i.e. VALUE IS SIGNIFICANT AT 0.005 LEVEL.

MEDIAN TEST: TYPE 3 ANSWERS IN THE ROULETTE AND MODIFIED ROULETTE EXPERIMENTS.

$$\text{MEDIAN} = \frac{58+6}{120} = \frac{64}{120}$$

∴ SCORES OF 1 OR MORE ARE ABOVE MEDIAN.

	ROULETTE EXPERIMENT	MODIFIED ROULETTE EXPERIMT	TOTAL
NUMBER OF SCORES ABOVE COMBINED MEDIAN	19	3	22
NUMBER OF SCORES BELOW COMBINED MEDIAN.	41	57	98
TOTAL	60	60	120

$$\chi^2 = \frac{120\left[(19\times 57 - 3\times 41) - \frac{120}{2}\right]^2}{22\times 98\times 60\times 60} = \frac{(1083-123-60)^2}{6468}$$

$$\therefore \chi^2 = 125$$

SIGNIFICANT AT 0·001 LEVEL (D.F.=1)

PASS SCORES FOR SUBJECTS AGED SIX TO TEN YEARS FROM THE FIRST BEADS AND ROULETTE EXPERIMENTS, TO PROVIDE DATA FOR MEDIAN TEST.

FIRST BEADS EXPERIMENT, PROBLEMS B1 - B4.

SUBJECT	NUMBER OF 'PASS' CHOICES	SUBJECT	NUMBER OF 'PASS' CHOICES
SB13	1	SB44	7
SB14	5	SB45	8
SB15	2	SB46	5
SB16	6	SB47	8
SB17	4	SB48	8
SB18	4	SB49	2
SB19	4	SB50	6
SB20	5	SB51	4
SB21	5	SB52	5
SB22	4	SB53	5
SB23	5	SB54	8
SB24	5	SB55	4
SB25	4	SB56	5
SB26	4	SB57	6
SB27	3	SB58	6
SB28	4	SB59	6
SB29	5	SB60	7
SB30	5	SB61	5
SB31	5	SB62	4
SB32	5	SB63	5
SB33	5	SB64	4
SB34	4	SB65	5
SB35	6	SB66	6
SB36	5	SB67	8
SB37	5	SB68	8
SB38	6	SB69	8
SB39	5	SB70	8
SB40	5	SB71	7
SB41	4	SB72	8
SB42	4	TOTAL	313
SB43	4		

ROULETTE EXPERIMENT, PROBLEMS R2-R5.

SUBJECT	NUMBER OF 'PASS' CHOICES	SUBJECT	NUMBER OF 'PASS' CHOICES
SR1	6	SR32	6
SR2	4	SR33	4
SR3	2	SR34	6
SR4	2	SR35	4
SR5	2	SR36	4
SR6	2	SR37	4
SR7	3	SR38	3
SR8	4	SR39	3
SR9	2	SR40	4
SR10	4	SR41	5
SR11	3	SR42	5
SR12	4	SR43	7
SR13	0	SR44	8
SR14	4	SR45	6
SR15	1	SR46	4
SR16	1	SR47	8
SR17	2	SR48	8
SR18	3	SR49	5
SR19	4	SR50	5
SR20	2	SR51	7
SR21	2	SR52	3
SR22	4	SR53	6
SR23	7	SR54	5
SR24	6	SR55	6
SR25	4	SR56	5
SR26	3	SR57	4
SR27	4	SR58	6
SR28	0	SR59	8
SR29	4	SR60	8
SR30	4	TOTAL	255
SR31	6		

MEDIAN TEST: PASS SCORES IN THE FIRST BEADS AND ROULETTE EXPERIMENTS.

$$\text{MEDIAN} = \frac{313+255}{120} = \frac{568}{120}$$

$\therefore$ SCORES OF 5 OR MORE ARE ABOVE MEDIAN.

	FIRST BEADS EXPERIMENT	ROULETTE EXPERIMENT	TOTAL.
NUMBER OF SCORES ABOVE COMBINED MEDIAN.	41	23	64
NUMBER OF SCORES BELOW COMBINED MEDIAN.	19	37	56
TOTAL	60	60	120

$$\chi^2 = \frac{120\left[(41\times37 - 23\times19) - \frac{120}{2}\right]^2}{64\times56\times60\times60} = \frac{(1517-437-60)^2}{107520}$$

$\therefore$ $\chi^2 = 9.68$.

PROBABILITY OF OCCURRENCE UNDER $H_0 < \frac{1}{2}(0.01)$ (ONE-TAILED TEST, D.F. = 1)
i.e. VALUE IS SIGNIFICANT AT 0.005 LEVEL

References

Arnheim, R. (1969). Visual Thinking. Berkeley, Calif.: Univ. of California Press.

Binet, A. (1890). La perception des longueurs et des nombres chez l'enfant. Revue Philosophique., 30, 68-81. (Trans. in Pollack, R.H. and Bremner, M.J., eds. (1969). The Experimental Psychology of Alfred Binet. New York: Springer).

Braine, M.D.S. (1959). The ontogeny of certain logical operations: Piaget's formulations examined by non-verbal methods. Psychol. Monogr., 73, 5.

Braine, M.D.S. (1962). Piaget on reasoning; a methodological critique and alternative proposals. In: Kessen, W. and Kuhlman, C., eds. Thought in the young child. Child Dev. Monogr., 27, 2, 41-61.

Brainerd, C.J. (1973 a). Judgements and explanations as criteria for the presence of cognitive structures. Psychol. Bull., 79, 172-179.

Brainerd, C.J. (1973 b). The origins of number concepts. Scientific American, 228, March, 101-109.

Brainerd, C.J. (1974). Postmortem on judgements, explanations, and Piagetian cognitive structures. Psychol. Bull., 81, 70-71

Brimer, M.A. (1967). The Classification of Qualitative Data. Unpublished document. University of Bristol, School of Education.

Brimer, M.A. (1973). A Technique for Classifying Qualitative Data. Unpublished document. University of Bristol, School of Education.

Bruner, J.S. (1966). On Cognitive Growth, I and II. In: Bruner J.S., Olver, R.R., and Greenfield, P.M. (1966). Studies in Cognitive Growth. New Yok: Wiley.

Bruner, J.S., Goodnow, J.J. and Austin, G. (1956). A Study of Thinking. New York: Wiley.

Bruner, J.S., Olver, R.R. and Greenfield, P.M. (1966). Studies in Cognitive Growth. New York: Wiley.

Bryant, P.E. (1971 a). Cognitive development. British Medical Bulletin, 27 No. 3 Supplement on cognitive psychology, 200-205.

Bryant, P.E. (1971 b). The young child's understanding of quantity. Paper delivered to the British Association, Swansea.

Bryant, P.E. and Trabasso, T. (1971). Transitive inferences and memory in young children. Nature, 232, 456-458.

Bryant, P.E. (1974). Perception and Understanding in Young Children., LondonMethuen.

References

Campbell, R. and Wales, R. (1970). The study of language acquisition. In: Lyons, J., ed. (1970). New Horizons in Linguistics. Harmondsworth: Penguin.

Carlson, J.S. (1969). Children's probability judgements as related to age, intelligence, socio-economic level and sex. Human Development, 12, 3, 192-203.

Carlson, J.S. (1970). The development of probabilistic thinking in children: a comparison of two methods of assessment. J. Genetic Psy., 116, 2, 263-269.

Carlson, J.S. and McMillan, D. (1970). Comparison of probability judgements between EMR and non-retarded children. Am. J. Mental Deficiency, 74, 5, 697-700.

Cellérier, G. (1972). Information processing tendencies in recent experiments in cognitive learning-theoretical implications. In: Farnham - Diggory, S., ed. (1972) Information Processing in Children. New York: Academic Press.

Cicourel, A.V. (1973). Cognitive Sociology. Harmondsworth: Penguin.

Chomsky, N. (1965). Aspects of the Theory of Syntax., Cambridge, Mass: MIT Press.

Cohen, J. (1957). Subjective probability. Scientific American, Nov. 1957. Reprinted in Messick, D. Med. (1968). Mathematical Thinking in Behavioural Sciences. San Francisco: Freeman.

Cohen, J. (1960). Chance, Skill, and Luck: The Psychology of Guessing and Gambling. Harmondsworth: Penguin.

Cohen, J. (1964). Behaviour in Uncertainty. Allen and Unwin: London.

Cohen, J., and Chesnick, E I. (1970). The doctrine of psychological chances. Brit. J. Psy., 61, 3, 323-334.

Cohen, J. and Christensen, I. (1970). Information and Choice. Edinburgh: Oliver and Boyd.

Cooper, E.S., Costello, A.J., Douglas, J.W.B., Ingleby, J.D. and Turner, R.K. (1974). Direct observation? Bull. Br. Psy. Soc., 27, 94, 3-9.

Davies, C.M. (1965). Development of the probability concept in children. Child Development, 36, 779-788.

Derks, P.L., and Paclisanu, M. (1967). Simple strategies and binary prediction by children and adults. J. exp. Psychol., 73, 278-285.

Dienes, Z.P. (1959). Concept Formation and Personality. Leicester: University Press.

Donaldson, M.C., and Balfour, G. (1968). Less is more: a study of language comprehension in children. Brit. J. Psychol., 59, 461-472.

References

Donaldson, M.C., and Wales, R.J. (1970). On the acquisition of some relational terms. In: Hayes, J.R., ed. (1970) Cognition and the Development of Language. New York: Wiley.

Fischbein, E., Pampu, I., and Manzat, I. (1970). Comparison of ratios and the chance concept in children. Child Development, 41, 377-389.

Flavell, J.H. (1963). The Developmental Psychology of Jean Piaget. New York: Van Nostrand.

Flavell, J.H. (1970 a). Concept development. In: Mussen, P.H., ed. (1970) Carmichael's Manual of Child Psychology. New York: Wiley.

Flavell, J.H. (1970 b). Developmental studies of mediated memory. In: Reese, H.W. and Lipsitt, L.P., eds. (1970). Advances in Child Development and Behaviour, Vol. 5. New York: Academic Press.

Flavell, J.H. (1971). Stage-related properties of cognitive development. Cognitive Psychol., 2, 421-453.

Flavell, .H. (1972). An analysis of cognitive developmental sequences. Genetic Psychol. Monogr., 86, 279-350.

Fraser, C., Bellugi, U. and Brown, R. (1963). Control of grammar in imitation, comprehension and production. J. verb. Learn. verb. Behav., 2, 121-135.

Furth, H.G. (1969). Piaget and Knowledge: Theoretical Foundations. New Jersey: Prentice-Hall.

Gelman, R. (1972 a). Logical capacity of very young children: number invariance rules. Child Development, 43, 75-90.

Gelman, R. (1972 b). The nature and development of early number concepts. In: Reese, H.W., ed. (1972). Advances in Child Development and Behaviour, Vol. 7. New York: Academic Press.

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

Ginsburg, H. and Rapaport, A. (1967). Children's estimates of proportions. Child Development, 38, 205-212.

Goffman, E. (1967). Where the action is. In: Goffman, E. (1967). Interaction Ritual. Harmondsworth: Penguin.

Goldberg, S. (1966). Probability judgements by preschool children: task conditions and performance. Child Development, 37, 157-168

Goodnow, J.J. (1972). Rules and repertoires, rituals and tricks of the trade: social and informational aspects to cognitive and representational development. In:

References

Farnham-Diggory, S., ed. (1972). Information Processing in Children. New York: Academic Press.

Gregg, L.W. and Simon, H.A. (1967). Process models and stochastic theories of simple concept formation. J. Mathematical Psychol., 4, 246-276.

Harré, R. (1971). Joynson's dilemma. Bull. Br. Psychol. Soc., 24, 83 115-119.

Harris, M. (1964). The Nature of Cultural Things. New York: Random House.

Hayes, J.R. (1972). The child's conception of the experimenter. In: Farnham-Diggory, S. ed. (1972). Information Processing in Children. New York: Academic Press.

Hecox, K.H. and Hagen, J.W. (1971). Estimates and estimate-based inferences in young children. J. exp. Child Psychol., 11, 106-123.

Hockett, C.D., (1960). The origin of speech. Scientific Amer., 203, No. 3, 88-96.

Hoemann, H.W. and Ross, B.M. (1971). Children's understanding of probability concepts. Child Development, 42., 221-236.

Hutt, S.J. and Hutt, C. (1970). Direct Observation and Measurement of Behaviour. Springfield, Ill: Thomas.

Inhelder, B. (1962). Some aspects of Piaget's genetic approach to cognition. In: Kessen, W., and Kuhlman, C., eds. (1962). Thought in the Young Child. Monogr. Soc. Res. in Child Dev., 27, No. 2.

Inhelder, B. (1972). Information processing tendencies in recent experiments in cognitive learning-empirical studies. In: Farnham-Diggory, S., ed. (1972). Information Processing in Children. New York: Academic Press.

Inhelder, B. and Sinclair, H. (1969). Learning cognitive structures. In: Mussen, P., Langer, J., and Covington, M., eds. (969). Trends and Issues in Developmental Psychology. New York: Holt., Rinehart and Winston.

Kelly, G.A. (1955). The Psychology of Personal Constructs. Vols. 1. and 2. New York: Norton.

Kelly, G.A. (1963). A Theory of Personality. New York: Norton.

Klahr, D. (1973). Quantification processes. In: Chase, W.G., ed. (1973). Visual Information Processing. New York: Academic Press.

Klahr, D. and Wallace, J.G. (1970a). An information processing analysis of some Piagetian experimental tasks. Cognitive Psychology, 1, 358-387.

References

Klahr, D. and Wallace, J.G. (1970b). The development of serial completion strategies: an information processing analysis. Brit. J. Psychol., 61, 243-257.

Klahr, D. and Wallace, J.G. (1972). Class-inclusion process. In: Farnham-Diggory, S., ed. (1972). Information Processing in Children. New York: Academic Press.

Klahr, D. and Wallace, J.G. (1973). The role of quantification operators in the development of conservation of quantity. Cognitive Psychology, 4, 301-327.

Kuhn, T.S. (1970). The Structure of Scientific Revolutions. Second Edition. Chicago: University Press.

Lowe, G. and Ranyard, R.H. (1973). Intuitive statistical judgements in children: estimates of proportion. Paper presented to British Psychol. Soc. Annual Conference, Liverpool.

Lunzer, E.A. and Pumfrey, P.D. (1966). Understanding proportionality. Mathematics Teaching, 34, 7-12.

Luria, A.R. and Yudovich, F. 1a. (1959). Speech and the Development of Mental Processes in the Child. Harmondsworth: Penguin.

Miller, G.A. (1956a). Information and memory. Scientific Amer. 195, 2, 42-46.

Miller, G.A. (1956b). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev., 63, 81-97.

Miller, G.A. (1964). Mathematics and Psychology. New York: Wiley.

Miller, G.A., Galanter, E. and Pribram, K.H. (1960). Plans and the Structure of Behaviour. New York: Holt, Rinehart and Winston.

Neisser, U. (1967). Cognitive Psychology. New York: Appleton-Century Crofts.

Nelson, K. (1973). Structure and Strategy in learning to talk. Monogr. Soc. Res. in Child Dev., No. 149, 38.

Newell, A. (1972). A note on process-structure distinctions in developmental psychology. In Farnham-Diggory, S., ed. (1972). Information Processing in Children. New York: Academic Press.

Newell, A. and Simon, H.A. (19 1). Computer simulation of human thinking. Science, 134, 2011-2017.

Newell, A. and Simon, H.A. (1972). Human Problem Solving. New York: Prentice-Hall.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. Acta Psychologica, 32, 301-345.

References

Pascual-Leone, J. and Smith, J. (1969). The encoding and decoding of symbols by children: a new experimental paradigm and a neo-Piagetian model. J. exp. Child Psychol., 8, 328-355.

Peterson, C.R. and Beach, L.R. (1967). Man as an intuitive statistician. Psychol. Bull., 68, 29-46.

Piaget, J. (1932). Language and Thought of the Child. London: Routledge.

Piaget, J. (1951). Play, Dreams and Imitation in Childhood. London: Routledge

Piaget, J. (1952). The Child's conception of Number. London: Routledge.

Piaget, J. (1954). The Construction of Reality in the Child. New York: Basic Books.

Piaget, J. (1960). In: Tanner, J.M. and Inhelder, B., eds. (1960). Discussions on Child Development, IV, 3-27, 77-83. London: Tavistock

Piaget, J. (1967). Biologie et Connaissance. Paris: Gallimard.

Piaget, J. and Inhelder, B. (1951). La Genese de l'Idée de Hasard Chez l'Enfant. Paris: Presses Universitaires de France.

Piaget J. and Inhelder, B. (1966). The Psychology of the Child. London: Routledge.

Piaget, J. Inhelder, B. and Szeminska, A. (1960). The Child's Conception of Geometry. New York: Basic Books.

Pinard, A. and Laurendeau, M. (1969). "Stage" in Piaget's cognitive-developmental theory: exegesis of a concept. In: Elkind, D. and Flavell, J.H., eds. (1969). Studies in Cognitive Development. Oxford: University Press.

Pire, G. (1958). Notion du hasard et développement intellectual. Enfance, 1958, 131-143.

Popper, K.R. (1959). The Logic of Scientific Discovery. London: Hutchinson

Pumfrey, P.D. (1968). The growth of the schema of proportionality. Brit. J. Educ. Psychol. 38, 2, 202-204.

Pylyshyn, Z. (1972). Are there discovery procedures? Bull Brit. Psychol. Soc., 25, 88, 189-192.

Reese, H.W. and Schack, M.L. (1973). Comment on Brainerd's criteria for cognitive structures, Psychol. Bull., 81, 67-69.

Reitman, W.R. (1965). Cognition and Thought. New York: Wiley.

Ross, B.M. (1966). Probability concepts in deaf and hearing children. Child Development, 37, 917-927.

Satterly, D.J. and Brimer, M.A. (1971). Cognitive styles and school learning. Brit. J. Educ. Psychol., 41, 3, 294-303.

Simon, H.A. (1972). On the development of the processor. In: Farnham-Diggory, S., ed. (1972). Information Processing in Children. New York: Academic Press.

References

Simon, H.A., and Siklossy, L. eds. (1972). Representation and Meaning: Experiments with Information Processing Systems. Englewood Cliffs: Prentice-Hall.

Sinclair-de-Zwart, H. (1969). Development psycholinguistics. In: Elkind. D. and Flavell, J.H., eds. (1969). Studies in Cognitive Development. Oxford: University Press.

Skinner, B.F. (1953). Science and Human Behaviour. New York: Macmillan.

Smedslund, J. (1964). Concrete reasoning: a study of intellectual development. Child Dev. Monogr., 29, 2.

Smedslund, J. (1968). Conservation and resistance to extinction: a comment on Hall and Simpson's article. Merrill-Palmer Quarterly, 14, 211-214.

Stevenson, H.W. (1970). Learning in children. In: Mussen, P., ed. (1970). Carmichael's Manual of Child Psychology. New York: Wiley.

Sullivan, F.J. and Ross, B.M. (1970). What is learned in probability learning? Developmental Psychology, 2, 1, 58-65.

Turiel, E. (1969). Developmental processes in the child's moral thinking. In: Mussen, P., Langer, J. and Covington, M., eds., (1969). Trends and Issues in Developmental Psychology. New York: Holt, Rinehart and Winston.

Turing, A. (1950). Computing machinery and intelligence. Mind, 59, 433-450.

Vygotsky, L.S. (1962). Thought and Language. New York: Wiley.

Wales, R. (1971). Comparing and Contrasting. In: Morton, J., ed. (1971). Biological and Social Factors in Psycholinguistics. Logos Press.

Wallace, J.G. (1972a). Stages and Transition in Conceptual Development. Slough: NFER.

Wallace, J.G. (1972b). The development of conservation of discontinuous quantity in the light of recent experimental findings. Appendix B in: Wallace, J.G. (1972). Stages and Transition in Conceptual Development. Slough: NFER.

Wallace, J.G., (1973). Class inclusion performance in children: information processing theories and experimental studies. Paper presented to British Psychol. Soc. Annual Conference, Liverpool.

Weir, M.W. (1964). Developmental changes in problem solving strategies. Psychol. Rev., 71, 473-490.

Werner, H. (1948). Comparative Psychology of Mental Development. Chicago: Follett.

References

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In: Harris, D., ed. (1957). The Concept of Development: An Issue in the Study of Human Behaviour. Minneapolis: University of Minnesota Press.

Werner, H. and Kaplan, B. (1963). Symbol Formation. New York, Wiley.

Wohlwill, J.F. (1963). Piaget's system as a source of empirical research. Merrill-Palmer Quarterly, 9, 253-262.

Wohlwill, J.F. (1968). Responses to class-inclusion questions for verbally and pictorially presented items. Child Development, 39, 449-465.

Woodworth, R.S. and Schlosberg, H. (1954). Experimental Psychology. Revised edition. New York: Holt, Rinehart and Winston.

Yost, P.A., Siegel, A.E. and McMichael, J. (1961). Non-verbal probability judgements by young children. Penn. State Univer., Res. Bull. No. 16. Described in Flavell, J. H. (1963), p. 393.

Yost, P.A., Siegel, A.E. and Andrews, J.M. (1962). Non-verbal probability judgements by young children. Child Development, 33 769-780.