STAGES IN PROBLEM SOLVING

A thesis submitted in partial
fulfillment of the requirements for the
Degree of Doctor of Philosophy

by

Hazel Eleanor Steiger

Australian National University

September 1970

This thesis describes original
research carried out by the author in the
Department of Psychology of the
Australian National University
from January 1968 to July 1970.

Hazel Steiger

ACKNOWLEDGEMENTS

It is a pleasure to acknowledge the help of my supervisor, Associate Professor G. Seagrim who, in the midst of his myriad other responsibilities, was able to devote conscientious attention to the manuscript in draft form.

I am indebted to Dr. Michael Cook for his suggestions in the early stages of the theoretical and experimental design of this thesis, but especially for his technical advice and assistance in the computer programming relating to the Monte Carlo simulations. I am extremely grateful to Paul Moens for the many hours he spent to ensure that these programs were in working order.

I thank Professor P.A.P. Moran and Dr. Martin Ward for their statistical and mathematical advice. Robin Milne was extremely helpful in his comments on a draft chapter, and many of his suggestions are incorporated here.

I am grateful to Dr. William Steiger who originally brought to my attention many of the questions which I have tried to answer in this thesis.

I am especially grateful to Michael Kahan for providing me with stimulation and encouragement by his trenchant criticisms and invaluable suggestions.

Finally, I am indebted to the Australian National University for providing me with a scholarship which made it possible for me to embark on this doctoral thesis.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1

THEORETICAL BACKGROUND

The history of problem solving has, with the exception of the Gestalt school (Köhler, 1925; Wertheimer, 1945; Duncker, 1945), been largely created by those psychologists whose interest lay in the manipulation of one or several independent variables for the purpose of observing how this affected one or more dependent variables. Subjects were studied in strictly controlled laboratory situations, as they solved what were often restricted and unchallenging problems. (Two noteworthy exceptions to this are provided by the work of Moore and Anderson, 1954, and John, 1957. Comprehensive reviews of recent problem-solving research and theory is presented in D.M. Johnson, 1955; Gagné, 1959; Duncan, 1959; G. Davis, 1966 and Kleinmuntz, 1966.) The object of this type of research was to express the performance of the problem solver in terms of number of solutions, number of errors, time to solution and so on. This orientation derives, in part, from a view of problem solving as closely related to learning, so that complex behaviour is perceived as not qualitatively different from simple discrimination learning, both being built up from simple S-R units. This is exemplified by articles by Skinner and by Staats in

Kleinmuntz (1966, pp.225-257; pp.259-339).

One cannot deny the importance of accumulating a body of such results by strict experimental methods, but it becomes necessary to counterbalance the sparse picture that emerges from a preponderance of this 'statistical' approach. Studies of the conditions under which problems are solved throw little light on the process of problem solving. The alternative approach suggested here is to forego questions about, say, the number of subjects with correct or incorrect solutions, or at least to supplement this by asking what kind of program, or hierarchical organization of activities would display the kind of behaviour observed in problem-solving experiments. By attempting to identify the content of behaviour, we avoid the danger of overlooking elements of that behaviour that may be hidden in or by the statistics.

Examination of the problem-solving literature reveals that, with a few exceptions such as Kohler (1925) and Maier (1930) little attention has been paid to the interaction of the solver with the problem itself. This thesis will, in part, address itself to this aspect of the situation. In general terms, the problem-solving situation is viewed here as one in which the solver interacts with an objective situation which contains its own requirements. By casting a single net around the solver and the problem, the emphasis

shifts away from the products of problem solving to focus instead on the process. An analysis of the problem, made in either formal or structural terms, enables us to clarify how the task is mapped into an internal representation to be handled by the solver's cognitive apparatus. A complete description of a subject solving a problem must include the quantitative and qualitative attributes of his behaviour as well as a thorough description of his task environment.

This work receives its direction from two sources: the formal framework is provided by the stages model[1] developed by Restle and Davis (1962) in which the distribution of solution times depends on the number of independent stages (k) of the problem-solving process as well as on the probability of passing a given stage (λ). The model used to describe the distribution of solution times is the gamma distribution. The methods and concepts of information-processing theory provide the tools to evaluate the psychological credibility of this statistical model. Both these

---

[1] It is worth noting that the stages model incorporated into this thesis is, as will be seen, qualitatively different in both intent and expression from the traditional stages models perpetuated by introductory psychology texts, in which there are said to be four stages: preparation, incubation, inspiration and verification. Although a more sophisticated model has been proposed (cited by Green, 1966), models which attempt a sequential description of functions in problem solving are unproductive, and contribute little, if anything, to a detailed understanding of the process itself. (This issue is discussed by several contributors in Kleinmuntz, 1966.)

areas will now be examined in greater detail - the stages model in section 1.1, and information-processing theories in section 1.3.

1.1 The Stages Model

The original, unpublished work in this area was done by J.H. Davis (1961) and reported by Restle (1961, 1962). Restle and Davis (1962) applied the statistical theory of waiting times to describe group and individual solution times on multistage problems. The theory states that the time to solution of a single stage is an exponentially distributed random variable. In the multistage case, the time to complete the whole problem is taken as the sum of the exponentially distributed random variables and is described by the gamma distribution. The distribution depends on the number of stages (k) and their probability density (λ) which is assumed to be constant for each stage. The number of stages in a problem is considered to be the number of distinct ideas or parts necessary to solve the problem. Each stage is passed with a certain probability per unit time. The model assumes that each stage is equally difficult, and that each subject has the same probability of passing or solving a given stage.

The number of stages in a problem, as well as the

probability rate are estimated from the data using the method of moments (Mood, 1951). Restle and Davis used three different Eureka, or self-confirming, word problems. 178 undergraduate subjects were asked to estimate the number of stages in each problem, and these mean estimates showed gross agreement with the statistical estimates. Each of the obtained distributions of solution times were satisfactorily fitted by the appropriate gamma distribution. The authors were interested in testing problems that have differing numbers of stages. They suggest that a taxonomy of problems is possible, using number of stages as a classifying principle. The paper then argues the relative merits of Equalitarian and Hierarchical theories of group problem-solving, and goes on to use proportions of solvers from the individual data as a basis for prediction and for generating a distribution of group problem-solving times according to waiting times theory. This aspect falls outside the domain of this study. A subsequent publication by Davis and Restle (1963) enlarges on the previous work, the main emphasis is on group behaviour.

J.H. Davis (1964) extended the evidence for the model by introducing four new problems, including a non-Eureka decoding task (Lorge, Fox, Davitz and Brenner, 1958). The experiments were designed to investigate whether a relationship existed between the number of stages as indicated by

logical analysis of the structural components of the problem, and that given by statistical estimation. The four problems were combined into two compounds of two, and the effect on solution times of these compound problems was investigated. The gamma function gave a good fit to the data. The total number of stages for a compound, estimated from the solution times data, seemed to be a product rather than a linear combination of the stages of the two component problems that comprised the compound. (The statistical theory of waiting times that underlies the stages model is discussed in greater detail in Chapter 2.)

The work done by both Restle and Davis has reached an impasse that is indicated by their statement that:

> Experimental separation of stages is needed before conclusions can be drawn about the exact nature of a stage. The necessary data was not collected in the present experiment. (1963, p.115)

In order to examine the assumptions of the model, and to isolate an individual stage, it was imperative to extend the information beyond that provided by the solution times, by discovering what the subject was doing while he worked on a problem. The appropriate methodology for this was suggested by research being carried out in the area of computer simulation of cognitive processes. We will return to a discussion of this methodology later in the chapter (1.4).

1.2 Information-Processing Theory

The computer simulation of human cognitive processes began as a research strategy in about 1958 with the work of Newell, Shaw and Simon, and forms an area within the larger field of artificial intelligence. The goal of artificial intelligence research is to produce computer programs that are capable of 'intelligent' behaviour (Feigenbaum and Feldman, 1963, Part I), whereas simulation research aims for programs that will behave in the same way as human beings. Many of the problems that preoccupy investigators of cognitive processes concern the status of computer programs as statements of theory, and the methods for validating or assessing the computer output in relation to the human output it is intended to model. It is not our intention to elaborate this further, but a large volume of literature is concerned with these questions (Newell, 1962; Neisser, 1963; Reitman, 1964; Frijda, 1967; Dennett, 1968). By viewing the problem solver as a processor of information, this work shares a common orientation with that of the computer simulation theorists.

With their emphasis on the process rather than the products of problem solving, and in order to guarantee a rich yield of data, information-processing theorists began to use far more complex problems than those usually reported in the psychological literature. Programs were written that

simulated the behaviour of people playing chess (de Groot, 1965) or checkers (Samuel, 1959), proving geometry theorems (Gelernter, 1963) and solving algebra word problems (Paige and Simon, 1966). The pioneering work in the field, however, was the General Problem Solver (GPS) of Newell, Shaw and Simon, which was first reported in 1958. The research has been developed into a theory which has had a profound effect on current cognitive research (Miller, Galanter and Pribram, 1960; Feigenbaum and Feldman, 1963; Reitman, 1966; Kleinmuntz, 1966).

Since a digital computer is not peculiarly a device for manipulating numerical symbols but a general purpose device capable of manipulating all kinds of symbols, it can be used as a model for human information processing. The basic premise of the theory is that complex thinking is built up of elementary symbol manipulation processes. More specifically, it postulates (1) a control system or memory which contains symbolized information that is stored hierarchically. No statements are made about the physical structure or other properties of the memory. (For a more recent development, see Reitman, 1965.) (2) There are a number of primitive information processes for operating on the information in memory. A list of primitive information processes might include: reading a symbol, recording a symbol, transferring a symbol, comparing two symbols etc. It is also

assumed that the system can take different courses of action, depending on the outcome of the comparing operation. (3) There is a precise set of rules for combining these processes into whole programs which provide an explanation, or a model, of the observable behaviour. The rules specify the order in which the processes are carried out.

The theory conceives of problem solving as the search through a set of potential solutions for a correct one. This view derives from Selz's 'anticipatory scheme' which postulates that the subject has a schema of the task, in which there exists a gap to be filled. Selz's theory is reviewed by Humphrey (1963, Ch.V.). An organized chain of mental operations is triggered off, each determined by the preceding operation and by the cues in the schema. de Groot (1965) incorporated Selz's search model and his conception of thinking as a series of linear operations in describing the behaviour of chess players.

Problem solving is not always adequately described by a search through a space of possible solutions. Miller, Galanter and Pribram (1960) point out that much of problem solving centers around the search for a method or a plan. A distinction must be made between a problem for which the subject has a method and which he only needs to apply, and a problem where the first operation consists of developing a

plan. This is the distinction made by the Greek mathematician Pappus (cited in Polya, 1957, pp.141-147) between a 'problem to prove' and a 'problem to find'. The former involves demonstrating that some assertion is either true or false, whereas the latter involves a search. Information-processing theories would include the search for a method as simply a more complicated kind of search: the search for a path between two points is a simpler version of a search for any other kind of unknown.

Search is neither as structured nor as orderly as the information-processing model would imply. On first being presented with a problem, the typical response of the subject is to evaluate the problem as a whole, to reformulate it, to speculate about the solution, rather than to proceed immediately on a systemmatic check of available alternatives. The typical subject does not consider all alternatives, but rather only some subset of them. In other words, he uses heuristic methods for structuring his search, and for selecting which subgoals to solve en route to the final solution.

1.3 Heuristics or Heuristic Methods

There is some disagreement as to how a heuristic or heuristic method should be defined. Although Polya uses the noun heuristic to refer to "the study (of) methods and

rules of discovery and invention" (1957, p.112),information-processing theorists use the term to denote any principle or device that contributes to the reduction in the average number of steps in the search to solution. Heuristics refer to any method or trick that will improve the efficiency of problem solving performance, by, for example, suggesting the order in which solutions should be examined, or by providing a cheap test which will distinguish likely and unlikely solutions. The other feature of heuristic methods is that they are seldom infallible: although they are often effective, it is impossible to guarantee that they will prove successful.

On the other hand, if a procedure exists for calculating a solution in a finite number of steps, or for ordering the search so that a solution is guaranteed after a finite number of trials, then that procedure is an algorithm. (Algorithms are a branch of modern mathematics, and an introduction to the theory is presented in M. Davis, 1958). For some problems, highly efficient algorithms are known, as, for example, finding the maximum of a function; however, in many problems either no algorithm exists, or those that do are excessively consuming of time and effort, such as exploring all possible continuations of a chess position or all possible letter-digit combinations in a cryptoarithmetic problem. It is at this point that heuristic methods provide

the most useful strategy.

Although earlier writers (Newell, Shaw and Simon, 1958a); Reitman, 1959) contrasted heuristic programs and algorithms, emphasizing the 'rule of thumb' character of the former, other work (Minsky, 1963) tends to lessen the polarization of the definitions so as to include algorithms as one kind of heuristic, or to distinguish between

> algorithmic properties of programs--limiting properties as the number of computations increases--and their heuristic power--their efficacy in solving problems with reasonable computing effort. (Newell and Simon, 1962,p.7).

Miller, Galanter and Pribram (1960, p.160) distinguish between heuristic and algorithmic 'plans' in terms of whether they are systemmatic or not, and, therefore, whether their effectiveness can be guaranteed.

However algorithms and heuristics are defined, the distinction is never simple, partly because the same exhaustive system can be either an algorithm or heuristic, depending on how the search region (the set of all possible solutions) is selected, in other words, whether it is finite or infinite. If the search region is infinite, then a systemmatic search through it would not be an algorithm. If the region is finite, however, then an exhaustive search would be successful, and the method would be an algorithm.

Further difficulties in precisely defining heuristics and algorithms arise since an algorithm is defined with reference to a defined class of problem, while a heuristic may be either general or specific. For example, a heuristic could be as general as reducing the size of the region in which search is to operate, or as specific as guessing only even numbers. An additional property of heuristics is that they operate at several levels in problem solving. Some heuristics, for example, act to impose limits on the search area, while others, such as guessing, act directly by organizing the search process itself (Kilpatrick, 1967). It is important to note that guessing is never random, since it is always channelled by some information about the relevant boundaries of the problem.

Not every imperfect method that appears to be a heuristic is actually one, nor is every heuristic an imperfect method. Further, a technique may be deceptively heuristic : for example, the checking of one's calculations at any point of the problem-solving process is not necessarily advantageous. If the solution is correct, checking it simply takes up time. However, if some error is discovered during this process, then checking is a heuristic. Instead of aiming for a rigorous definition of heuristic with its attendant difficulties, we will accept that used by Kilpatrick (1967),

> ...and define a heuristic as any device, technique, rule of thumb etc, that improves problem-solving performance. We consider heuristics to be typically provisional, without guarantee of effectiveness, but we do not attempt to contrast them with algorithms. If algorithms are heuristics, they are the least interesting sort for our purposes. (p.19)

The interest in heuristics is not restricted to current psychological research. Duncker (1945) obtained evidence that heuristic methods were used by his subjects. Duncker used complex problems, and the subjects were compelled to solve them in successive 'chunks' rather than all at once. He writes:

> The solution of a new problem typically takes place in successive phases which...have, in retrospect, the character of a solution, and in prospect that of a problem. (p.18)

In other words, problem solving consists of decomposing the problem into intermediate phases. The theory of productive thinking maintains that the subject analyzes both the situation he is in and the goal, locates a source of difficulty and then attempts to remove it. For Duncker, whereas solutions are means to a goal, heuristic methods are means to a solution.

Duncker's work, originally published in 1935, does not appear to have stimulated further research. Part of the renewed interest in heuristics has come from the work of the mathematician Polya who has written several volumes on the subject (1954; 1957; 1962-5). Part of this interest

is also due to the development of high speed computers which have made possible the simulation of human behaviour by machines. Programmers have discovered that, without the inclusion of heuristic rules into problem-solving programs, the machines either use up their storage capacity or require an excessive time to reach a solution. For example, Gelernter, Hansen and Loveland (1963) in developing a program that proved geometry theorems, discovered that by incorporating a single heuristic rule, the average number of alternatives considered by the machine at each stop was reduced from one thousand to about five.

Newell, Shaw and Simon (1958a) believe that a theory of problem solving is concerned with discovering and understanding systems of heuristics. GPS embodies two very general such systems: means-end analysis and planning.

Means-end analysis

The principle underlying means-end analysis is that progress occurs by substituting for the achievement of a goal the achievement of a set of subgoals. (Although the language is different, it will be seen that the principle is the same as that espoused by Duncker.) For every new difficulty that arises, a new subgoal is created to overcome this difficulty. The essence of means-end analysis consists in (1) matching two objects to find a difference D, (2) attaining a relevant operator Q that will reduce this

difference and (3) applying the operator Q. A subject, using this heuristic method, is assumed to go through a process that runs: "Is there any way of transforming the given situation (A) into the desired solution (B)? I do not know a way. See what the difference is. The difference is D. How can I reduce D? Operator Q will reduce D, but I do not see how to apply it. Transform A so that Q can be applied. Now apply Q. Get A'. Now the problem is to get from A' to B. I do not know a way. See what the difference is..."

We can observe three main goal types in this process:

Transform situation A into situation B

Reduce difference between situation A and situation B

Apply operator Q to situation A

The methods for each of the goal types are interrelated and recursive in organization. The general principle underlying the means-end heuristic is subgoal reduction, and the successful application depends on how shrewdly the measure of difference is defined by the subject and how appropriately the transformations are selected.

Planning

The second very general principle incorporated into GPS is planning, which, simply, consists in omitting certain features or details of the problem in order to simplify it.

The plan used for solving the simplified problem is then used as the strategy for solving the original, more complex problem. The following example from a problem in propositional calculus will clarify the point. By ignoring differences among logical connectives, and focussing only on the symbols and their groupings, logical operators that add, delete or regroup symbols can then be applied to the abstracted propositions, regardless of their connectives. If this simpler problem can be solved, possibly by using means-end analysis, the steps in its solution can suggest a plan for solving the original problem. Another example of planning entails dividing the original problem into a number of parts, each of which can be attacked by a smaller search procedure. This reduces search time not by a fraction, but by a fractional exponent. It is therefore strategic to expend considerable effort to discover such 'islands' in the solution of complex problems (Minsky, 1963).

Newell, Shaw and Simon observed the use of planning heuristics in their more proficient subjects (1958b). Gelernter (1963) included planning heuristics in planning the Artificial Geometer, a program which proves theorems in elementary plane geometry. The planning heuristic led to the drawing of a diagram, enabling relationships to be determined in the diagram, a simplified model of the original problem.

For Miller, Gallanter and Pribram (1960), a plan is more general, so that a Plan for an organism becomes synonymous with a program for a computer. Essentially their theory postulates that a hierarchical structure forms the basic organization of behaviour in general, and problem solving in particular, so that:

> A Plan is any hierarchical process in the organism that can control the order in which a sequence of operations is to be performed. (p.16)

The importance of planning heuristics is emphasized in the formation of plans, and the authors conclude:

> ...a plausible account of heuristic plans... will provide the general outline within which a theory of thinking about well defined problems can eventually be constructed. (1960, p.179)

We may summarize this section with Polya's statement:

> We have a plan when we know, or know at least in outline, which calculations, computations, or constructions we have to perform in order to obtain the unknown...The main achievement in the solution of a problem is to conceive the idea of a plan. (1957, p.8-9)

We have seen that the two general heuristics of means-end analysis and planning have been confirmed by several problem-solving investigators. As Miller, Galanter and Pribram (1960, p.167) point out, psychologists take heuristics for granted by failing to notice the multitude of alternatives open to the subject that he never tests simply because he never considers them, having automatically

rejected a great number of them as being impossible or unfruitful. Poincaré (1952), writing about mathematical discovery says:

> Discovery consists precisely in not constructing useless combinations, but in constructing those that are useful, which are an infinitely small minority. Discovery is discernment, selection. (pp.50-51)

Heuristics provide a way of applying this discernment, and the study of problem solving can seriously be considered the study of efficient search techniques.

1.4 The Protocol Method

The information-processing theorists, with their concern for the process of problem solving, have been responsible for a renaissance of interest in subjective report as a source of data. In this section, the protocol method, its advantages and limitations, will be discussed; this is followed, in the next section (1.5) by a review of studies examining the effects of verbalization on problem-solving performance.

The technique employed by information-processing and other investigators involves instructions to the subject to 'think aloud' as he works on a problem. The tape recording of the verbal behaviour is transcribed, and provides the material for analysis. This is a protocol. The method

yields a record of the time sequence as well as of the continuous verbal behaviour of each subject.

The protocol method is not a recent innovation. It was used by the Introspectionists at the beginning of the century, and Duncker (1945) depended on protocols to develop his theory of problem solving. de Groot, in the 40's, obtained his information about the thought processes of chess players by asking them to think aloud as they made their moves. (This work has only recently been translated into English, 1965.) With the advent of behaviourism, the method was cast into disrepute, not to be resurrected until the development of electronic computers and the subsequent interest in the simulation of human processes.

Although the technique of obtaining protocols is simple enough, it raises many serious questions, not all of which can be answered satisfactorily. The thinking aloud procedure provides a record of the subject's activity, but the record is not a complete one. The subject probably makes a decision (which may be subject to vacillation) concerning the level of his reporting so that he will not verbalize below it, but it cannot be assumed that he reports everything even at this level. A related difficulty is that much of the paralinguistic material in the recording is lost during transcription.

Another question arises around the accuracy of the report. How faithfully does verbalization reflect behaviour? The subject may report correctly on what he is doing, but he may also contradict what he is doing. What transformations do subjects make when speaking about their thoughts?

A possibly more serious danger is inherent in the process of verbalization itself. Although there is reason to believe that the process of verbalization is functionally parallel to the thinking process, it is also true that verbalization, chronologically, lags behind thought. How do verbalization and thinking interact with each other? How do strategies and decisions made during problem solving influence verbalization (Weinberg, 1965)? It may be that thinking inhibits speech, or it may be that the verbal process cannot be synchronized with the rapid and irregular nature of thought. Silence on the part of the subject may coincide with extremely organized mental activity, or it may be the counterpart of cognitive anarchy that cannot be verbally tethered. Alternatively, a temporary cessation in verbal activity might occur simultaneously with the subject experiencing 'a mind gone blank'.

A corollary of the feedback relationship between thinking and verbalizing is that the process of verbalization

may influence thought to transform it into a product that is quantitatively or qualitatively different from thought proceeding under normal conditions. It is possible, for example, that verbal facility aids problem solving, perhaps by activating symbolic processes. However, comparison of behaviour of subjects in a binary choice experiment did not reveal any major differences between subjects who did think aloud and those who did not (Feldman, 1963).

No claims are made for answers to these questions in these pages. Although some of the problems are tackled in this study, the intention is to counsel restraint in evaluating this method, and to invite future research aimed at perfecting a potent tool.

> We can ignore questions about the changes in thinking produced by vocalization, however, if we restrict our use of the method to producing hypotheses rather than to validating them. The dangers of obtaining incomplete or distorted information vanish if we insist that data obtained by asking subjects to think aloud be subjected to experimental test. The only alternative is to disallow evidence obtained by means of thinking aloud, and that seems much too high a price to pay. (Kilpatrick, 1967,p.8)

The next section deals with experimental evidence arising from various attempts to externalize thought.

1.5 Experimental Evidence of the Effects of Verbalization

Psychologists have been varied and ingenious in their attempts to discover 'what is going on in the head'. Some investigators, using what we may term the manipulative approach, have presented their subjects with physical problems such as wire puzzles (Ruger, 1910) jars of water (Luchins, 1942) or clamps and sticks (Maier's pendulum problem, 1930). Although manipulating objects may provide a manifestation of some of the hypotheses entertained by a subject, this method does not disclose all his hypotheses nor the ones he rejects.

A related method involves the use of materials that will be a direct portrayal of the choices made by a subject. Lazerte (1933) designed an envelope test in which the subject, at each step, selects one of several alternative procedures typed on different envelopes. When he has decided on a procedure, he opens the envelope and finds another set of envelopes with alternative procedures and so on, until a solution is arrived at. The experiment provides an objective description of the solution path followed by the subject. Buswell and Kersh (1956) modified this technique which has been widely used, particularly at the Loyola Psychometric Laboratory (Rimoldi, Fogliatto, Erdmann and Donnelly, 1964). This structured-choice situation produces

a very artificial task environment, since the subject is presented with a list of alternatives instead of having to generate his own. A method which places a high degree of constraint on the subject's behaviour ignores an essential aspect of problem solving: discerning the available alternatives and then generating decision structures to deal with them.

Other experimental approaches have involved the direct use of some form of verbal report as an index of unobservable processes. The first of these is the predictive method, since subjects are instructed to say what they are going to do before actually doing it. Ray (1957) used a simple manipulative problem that involved a panel of seven switches, any two of which, when depressed simultaneously, turned on a light, the required solution. The results showed that the 'verbal' group was more systemmatic in its approach, and had fewer errors and fewer trials to criterion than the 'nonverbal' group. Gagné and Smith (1962), using bright 14-15 year old boys, gave instructions to state aloud the reasons for each move in a series of 'Tower of Hanoi' problems. The authors conclude that the reason for the significantly shorter solution times for this group, and for the fewer moves to solution was that verbalization forced the subjects to think of new reasons for their moves,

thus facilitating the discovery of a general principle at the end of the task. However it is misleading to conclude from this that thinking aloud enhances performance. The results are ambiguous, since the instructions to verbalize are not distinguished from the requirement to justify each step. It is possible that writing down the reasons for each move would have led to the same result.

The Gagné and Smith study was replicated by Davis, Carey, Foxman and Tarr (1968), who hypothesized that it was not verbalization per se that was important, but that the enhanced performance was due to social facilitation arising from the presence of the experimenter. The experiment confirmed the Gagné and Smith results when verbalization occurred in the presence of the experimenter, although this varied with task demands. It is not clear from the report whether subjects verbalized as they worked on the problem, or whether they stated their reasons for moves or a sequence of moves just before enacting them.

Perhaps the earliest method used by psychologists to externalize thought was the method of introspection. Binet (1903) presented his subjects with problems and instructed them to report on their thinking as they worked on them. Although introspection did yield observable data, serious problems were raised about the character and extent of distortion produced by asking the subject to observe himself

as he thought. D.M. Johnson (1955) points out that there is substantial disagreement among subjects introspecting on the same problem.

It is clear that both introspection and retrospection result in reports that are far removed from the ongoing process itself, requiring as they do analyses of the structure of thought. Introspection demands that the introspector make himself the object of his attention, which effectively transforms his role into that of experimenter. If this is the case, we would expect that considerable practice would be necessary before the subject could effectively report on himself, and this would lead to the further danger of the skill being confounded with the process itself. The weakness of the retrospective method is the tendency of the subject to construct his mental activity so that it appears to be more structured and orderly than it actually was.

It is nevertheless possible to obtain a sequence of behaviour that will serve as the basis for analysing the process of thought by using a method which has none of the problems of the methods discussed above. By asking the subject to '<u>think aloud</u>' as he works, we ask him only to provide an account, not an analysis, of his thoughts. We do not run the risk of behaviouristic criticism, since the verbal utterances of the subject are just as much behaviour as his

eye movements or his patellar reflex. The subject is simply emitting a continuous stream of verbal behaviour.

The technique (réflexion parlée) appears to have originated with Claparède (1917) who was well aware of its advantages and disadvantages. Although the method requires no skill, some investigators have thought it necessary to train their subjects in thinking aloud (Patrick, 1934; Sargent, 1940).

The evidence on the effects of verbalizing is scanty as well as inconclusive. Hafner (1957), using two matched groups of 4th grade children, instructed the experimental group to verbalize as they worked on stencil tests. This group needed significantly fewer moves to solve the problem, but time to solution and number of solutions, although in the expected direction, were not significantly lower than for the nonverbal group.

Verbalization can affect the correlation between successful performance on two tests, as shown by Brunk, Collister, Swift and Stayton (1958). Subjects who were asked to verbalize on the second of two Vigotsky type concept attainment problems showed a lower degree of correlation between the problems than the nonverbal subjects. The results do not indicate whether verbalization aids or impedes performance.

Roth (1966) found no difference in the number of correct solutions, solution time and dominant mode of imagery (measured by a questionnaire) between subjects who verbalized and those who did not, on four reasoning problems. The small number of subjects used in each group in this experiment however, casts doubt on the reliability of these findings.

Verbalization does not affect the speed with which multidigit multiplication problems are solved. In their report, Dansereau and Gregg (1966) conclude that the two component effects of verbalization act to cancel each other out: although 'thinking aloud' leads to gains in memory, it slows down processing speed. These independent factors are hidden in experimental designs where the intention is simply to compare solution times of verbalizing and nonverbalizing groups, and a worthwhile contribution could be made by isolating these factors and investigating them independently. A differential effect of verbalization according to level of problem difficulty is reported by Nance and Sinnot (1963) who found that verbalization slowed performance on hard anagrams, but decreased the time to solve easy problems.

In an experiment by E.S. Johnson (1964), the subject was required to indicate his current state of knowledge to the experimenter at regular intervals, and this interaction

was found to facilitate the thinking aloud procedure. The experimenter's presence should be introduced with caution however, since his interference or participation could lead to as many impurities in the verbal record as introspection.

Verbalization can destroy a recently formed concept according to Phelan (1965). Arguing that the ability to verbalize a logical or rational concept is a poor criterion for concept attainment, he points to classificatory concepts that may be in a state inaccessible to verbalization. An unsuccessful attempt to verbalize a concept can destroy it, whereas appropriate or successful verbalization facilitates the future use of a concept. Incorrect verbalization may interfere with rules and guides to actions that might otherwise have functioned adequately. Palzere (1967) however, using secondary school students, found no significant differences between the problem solving ability of subjects who do not verbalize a concept after demonstrating awareness of it, and those who do verbalize. The critical factor seems to be the degree to which a subject has consolidated a concept.

Additional evidence for the facilitating influence of verbalization comes from an experiment by Bower and King (1967) in which subjects were required to verbalize their hypotheses before making classification responses in a bidimensional classification learning problem. Those subjects

who verbalized their hypotheses before obtaining feedback made significantly fewer errors than the others.

Although the evidence from verbalization experiments is inconclusive, and although the methods used to obtain verbal reports vary widely, we will venture some tentative generalizations as to the nature of the mechanism involved. It appears that verbalization acts to raise the vigilance of the problem solver, so that he needs a higher level of acceptability for a hypothesis when he is required to say it out loud. We have seen several experiments testify to the superior performance of the verbalizer. We cannot assume that the verbalizer actually entertains fewer false hypotheses--we can only say that he manifests fewer. Being forced to verbalize before the concept or hypothesis is consolidated, as in the Phelan experiment, has a destructive influence. It is conceivable that if the subject had been able to verbalize when he was ready to, as in most 'thinking aloud' experiments, then this negative effect would not have appeared. This is supported by other work (Feldman, 1963). Furthermore, it seems reasonable that verbalization should have a differential effect that varies with the level of difficulty of the problem being used: if verbalization constitutes an extra task for the subject, then it could operate as either an aid or a burden, according to the task demands of the problem situation.

We have seen that the 'thinking aloud' method is both easy and effective to use. We have seen its weaknesses, and we have noted doubts that the protocol constitutes the parallel but independent source of data we are seeking. We have also acknowledged that it would be unimaginative to dismiss the method entirely, and have resolved instead to develop means to improve our techniques. With these provisos in mind, it becomes clear that the problem we now face is what to do with the rich harvest of data we have collected.

A stimulating answer to this dilemna has been provided through recent work reported by Newell (1966a; 1967) in which he has developed a meta-language for protocol analysis which represents a stage towards the development of a program to simulate the solving of cryptoarithmetic problems.

1.6 Protocol Analysis (Newell)

This account is taken directly from Newell's two papers in which he deliberately restates information-processing theory so that it may be applied smoothly to data from several tasks, including chess, cryptoarithmetic, logic and the Missionaries and Cannibals puzzle.

It is necessary to describe his theoretical framework since the language for analysis of the protocol emerges directly from it. The theory also provides a complete, if

stark, statement of how problem solving proceeds, a description of which is intended to supplement the review of information-processing theory presented earlier (1.2 and 1.3). The theory assumes an information-processing system comprised of a memory of symbolic structures, organized as a network constructed of labelled associations between words (as compared to existing computers in which memory is organized in the form of a fixed array of words to be addressed numerically): a central processor, serially organized, which has access to and restructures the memory; and input-output structures.

The information-processing system is a universal machine that can carry out arbitrary symbolic processes. The question of how the system is structured is bypassed, and the theory deals instead with methods, organization of processes, information etc. that form the program used by a problem-solving system.

Problem solving occurs in a problem space, whose elements consist of a set of positions or nodes, each of which represent a state of knowledge about the problem. In an operational sense, these elements are data structures in the memory, and represent internally what is known about the task environment. The initial situation and the desired situation (goal) both form elements of the problem space.

Associated with the problem space are a set of operators which are applied to the elements of the space to produce new elements. It is through the application of operators that new information about the problem is extracted from the old.

Problem solving consists of search, from some initial position to some final position which includes the solution. Behaviour is partly determined by the individual: the way in which evaluations of new or former positions are made, the manner in which new operators are selected and applied are a function of the subject's intelligence.

Search in a problem space is constructive, since the solver has to generate the elements. They do not exist independently. Newell elaborates:

> In essence, problem spaces are always exponentially growing trees: two independent paths cannot end up at the same element of space. One cannot do in a problem space what one does in a forest: put marks on trees to recognize the same place if it is returned to. In the problem space, a data structure may be generated that is identical in strucutre and content to another--but it will not be the same data structure, hence will not contain any 'tree mark'. (1966a, p.8)

There are several strategies or methods according to which search may proceed. In the three examples that follow, the search strategies are position-dependent, that is, the search is directed with reference to a particular position

or node in the problem-solving process. In the depth-first strategy, once a position has been consolidated, it becomes the focus for further search, so that all positions that are generated from that position are searched before the system returns to any other positions. The breadth-first strategy dictates that all positions at the same depth as the present position are generated and searched before the system proceeds to a deeper level. The memory requirements for these two strategies are considerable, since they entail retention of all positions leading to the current position in the depth-first strategy, and retention of all positions at a given depth for the breadth-first strategy. Search strategies which involve repeated search over the same region place less strain on memory. An example of such a strategy is that of progressive-deepening, which in chess games entails that only the present position be retained, since search always returns to this position. An example of a position-independent strategy is one in which search proceeds according to the structure of the operators themselves, rather than with reference to the current position. An example of this is search that is directed according to an operator that generates and tests the values of digits in ascending order. Before the subject can employ any of these strategies however, he must create or select for

himself the space in which he is to solve the problem. Although he has been presented with an external representation of the problem, the information thus contained must be encoded into an internal representation, and this may be expressed in the same form as the original problem, or, using the example of algebra word problems, in algebraic form, or both; it is in this internal representation that the transformations required by the operators take place (see section on means-end analysis). The question of internal representation is analysed by Newell and Ernst (1965).

The problem-solver may operate in more than one space. If the first proves inadequate, he may generate a new one. Alternatively, he could operate simultaneously in more than one. The planning heuristic for example could provide one space, which may be abandoned after a while, or referred to from time to time during the path to solution. This is true in Gelernter's (1963) geometry theorem proving machine, where one space is provided by the symbolic expressions which represent theorems, and the other consists of coordinates which represent the diagram. In the human subject analysed by Newell (1967), there were at least two problem spaces: the one in which writing occurred (external space), and, since there were long periods in which no

writing occurred, another one (internal space), the subject presumably working back and forth between them. However, the total problem-solving system is more than just one or more problem spaces. Retrieval processes, constant information (say, the multiplication table) and the processes which take it into account, may not be represented in a problem space. This is not to deny their critical importance.

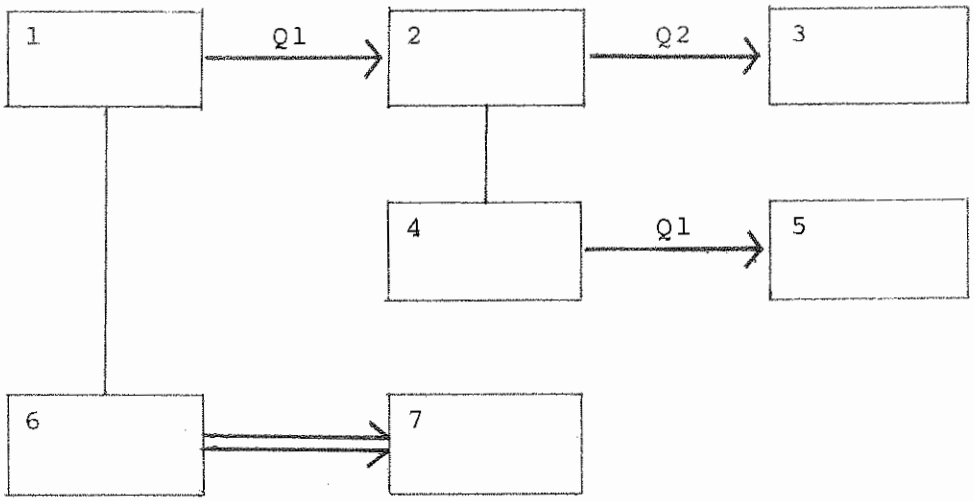
Newell's Meta-language

One cannot hope to do more than merely sketch Newell's method of protocol analysis: it is extremely detailed, proceeding to a micro-level well beyond the immediate interest of this thesis. In the later (1967) paper, 132 pages are required to analyze a twenty-minute protocol of a single subject. We will have reason to return to certain prominent features of his method in the later pages of this study (Chapter 7) where a far grosser level of analysis borrows key concepts from his papers, seeking to extend the stages model by applying to it appropriate information-processing concepts. Although this particular language for protocol analysis may serve as an orientation, we must remember that it was developed for a different purpose, and represents a tool much finer than is necessary for our objectives.

The information-processing theory described above is applied to the protocol through the Problem Behaviour Graph (PBG) which tracks the subject's search by (1) expressing the kind of information contained in the states of knowledge and (2) specifying operators such that each change in a state of knowledge corresponds to the application of one of the operators. The first level of analysis is concerned with describing search in this manner. No mention is made of the rules by which selections of operators, or decisions to apply them, are made.

The PBG was devised so that the total extent of the search, as well as its time history could be described. Any given node contains path information about how that node was reached, and past attempts information about previous actions while in this state. This is illustrated in Figure 1.1.

Analysis of the PBG reveals, for the cryptoarithmetic problem, four operators which are defined in terms of input-output characteristics, and which account for all transitions to a new state of knowledge. The operators act to process columns in the cryptoarithmetic sum, to generate values for a letter, to assign values to a letter and to test whether a certain value is admissible for a letter. In another problem, in a logic task for example, different



Rules for Problem Behaviour Graph (PBG)

- A state of knowledge is represented by a node (box).
- A horizontal arrow represents the application of an operator (Q) to a node; the new node thus produced is represented by a box at the head of the arrow.
- A return to the same state of knowledge as node X, is represented by another node below X, connected to it by a vertical line.
-A double horizontal line indicates repeated application of the same operator to the same state of knowledge.
- Time runs to the right and down.

The rules apply to the above diagram in the following way: Processing begins in Node 1. Operator Q1 is applied, producing the state of knowledge represented by Node 2. Operator Q2 is applied, leading to Node 3. In Node 4, the subject has returned to the same state of knowledge as in Node 2. However, this return was not effected by the application of an operator but through some other operation outside the problem space, as for example, recalling the earlier state. Q1 is applied again at Node 4, leading to Node 5. Node 6 represents a return to the state of knowledge contained in Node 1. Because it occurs later in time than Node 5, it appears below it. Q1 was applied again, as indicated by the double line.

FIGURE 1.1: The Problem Behaviour Graph (adapted slightly from Newell, 1966a).

operators would be generated. (The operators used in cryptoarithmetic tasks are discussed in greater detail in Chapter 7, where they are illustrated by examples from the subjects' protocols.)

If the subject's behaviour is to be explained by a sufficient set of processes, then there must be some description of how, *inter alia*, the four operators are selected, how positions are evaluated or lines of search terminated.

To do this, Newell embarks on the second stage of analysis, a language to express these decision and selection procedures. For this, he uses the *production system*, which has been widely used in logic and in the theory of algorithms (Markov, 1954). Given that the data can be construed as a set of correspondences between states of knowledge and the resultant actions, then the rule can be applied that

cues in knowledge state ——> action

or in more general terms,

condition ——> action

which is a *production*, and a set of productions constitutes a production system.

Newell found four types of productions in the protocol of the cryptoarithmetic subject. These are divided by function into those for (1) selecting an operator, (2)

setting up a goal, either to get something or to check something, (3) terminating a line of search, for example by finding that a certain conclusion is not possible and (4) repeating previous paths, for example to clarify a process.

The subject's behaviour is explained by

(1) the formalization of the problem space which includes the information contained in the state of knowledge, and operators which account for all transitions to new states of knowledge.

(2) the PBG which describes the states of knowledge, and plots the paths going through them from the initial to the desired state.

(3) the production system which explains some of the regularities in the subject's behaviour at each node of the PBG.

These constitute a sufficient set of explanations of the ongoing processes.

Our concern will be less with individual regularities than with whether the stages model is useful to describe general patterns in the behaviour of a group of problem solvers. We shall return to the information-processing level of discussion toward the end of this thesis, after observing the subjects from several angles, both real and simulated.

CHAPTER 2

THE WAITING TIMES MODEL

This chapter will present a detailed description of the statistical theory of waiting times which underlies the stages model used by Restle and Davis (1962) to describe solution times in problem solving. This will be followed by a short review of applications of the model in psychological research as well as in the representation of phenomena from other scientific fields. The implications and assumptions of the model are discussed, and suggestions for more realistic definitions are made. The contents of this chapter point to the need for simulation studies to provide a wider context in which to test the stages model.

2.1 The Theory of Waiting Times

(a) The discrete case : the geometric distribution

Suppose we perform a binomial experiment in which we are interested only in whether an event occurs (A) or does not occur ($\bar{A}$). Assume the experiment is performed over repeated trials, so that each trial is independent, and that on each trial

$$P(a)=p \text{ and } P(\bar{A})=1-p=q, \quad \text{where } 0<p<1$$

Suppose we repeat the experiment until the first occurrence

of A. The number of trials is therefore a random variable.

Define the random variable X as the number of trials necessary up to and including the first occurrence of A. X thus assumes the values 1,2... Since X=n if, and only if, the first (n-1) trials result in $\bar{A}$, while the nth trial results in A, we have the binomial probability

$$P(X=n)=q^{n-1}p \qquad \text{for } n=1,2... \qquad \text{(Eq.2.1)}$$

A random variable with probability distribution (Eq.2.1) is said to have the geometric distribution with parameter p.

Properties of the Geometric Distribution

(i) Graphical representation

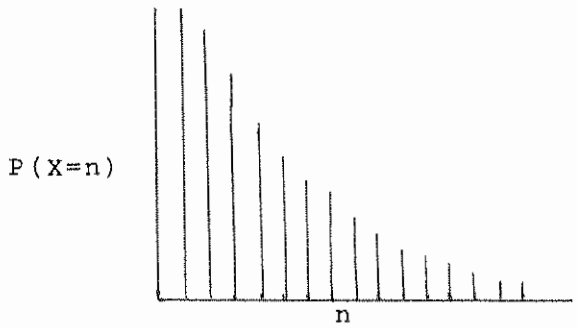


FIGURE 2.1

(ii) If the number of the trials is n, and assuming independence, we find that

$$P(1)=p$$
$$P(2)=p(q)$$
$$P(3)=p(q)^2$$
$$P(n)=p(q)^{n-1}$$

that is, at each trial the probability of the occurrence of event $\bar{A}$ is decreased by a fraction q of the probability of

the previous trial.

(iii) If X has a geometric distribution as given by (Eq.2.1), then $E(X)=1/p$ and $V(X)=q/p^2$.

Note: The fact that E(X) is the reciprocal of p has intuitive plausibility, since it says that the smaller the value of $P(A)=p$, the longer one must wait, on the average, for the first occurrence of event A.

(iv) The waiting time distribution for the first occurrence of an event can be used as a model for the following situation: imagine that a process occurs continuously in time, and is observed at discrete, uniformly spaced units of time. At each observation, a specified event of interest either has or has not occurred. There is no sense in which the event partially occurs. The waiting times model applies to the situation when the following waiting times assumption holds in the experiment:

(v) Waiting Times Assumption: At each observation (trial), given that the event has not occurred on a previous trial, there is a fixed, i.e. constant, conditional probability, p, that it will occur on the next trial. This means that, given the event has not occurred on trial 1, the probability that it will occur on trial 2 is the same as the probability that it will occur on trial 2000, given that it has not occurred by trial 1999.

In this sense the geometric distribution has 'no memory'. The information of no successes in the past is 'forgotten' so far as subsequent events are concerned. There is no contradication in saying that the response probability remains constant over all values of n, even though the ordinates of the probability distribution show a regular decay as n increases. The decay simply reflects the decreasing probability of getting long runs of $\bar{A}$ responses.

This argument will be extended to the exponential distribution which is a continuous analogue of, and has properties similar to the geometric distribution. A direct proof of this continuous approximation is found in Bush and Mosteller (1955, p.315-316).

(b) The discrete case : the negative binomial (Pascal) distribution

An obvious generalization to the geometric distribution arises in the case where we are interested in multiple occurrences of event A. Suppose an experiment is continued until event A occurs for the rth time. If

$$P(A)=p \text{ and } P(\bar{A})=q$$

on each trial, then the random variable X is defined as the number of trials necessary for the rth occurrence of A. Formerly we were interested in the first occurrence of A, but now we focus on the rth occurrence instead.

The event (X=n) is seen if, and only if, the first (n-1) trials resulted in precisely (r-1) successes, and the rth success occurred exactly on the nth trial. The probability of this event is simply $p\binom{n-1}{r-1}p^{r-1}q^{n-r}$, since what happens on the first (n-1) trials is independent of what happens on the nth trial.

Hence,

$$P(X=n) = \binom{n-1}{r-1}p^{r}\ q^{n-r} \quad \text{for } n=r,\ r+1,\ldots \text{(Eq.2.2)}$$

A random variable with probability distribution (Eq.2.2) is said to have the negative binomial distribution with parameters r and p. Note that when r=1, the negative binomial distribution reduces to the geometric distribution of (Eq. 2.2).

Properties of the Negative Binomial Distribution

(i) Graphical representation

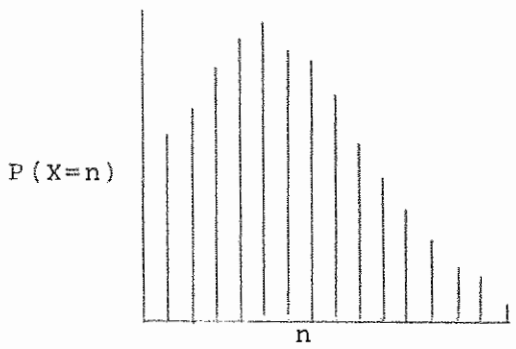


FIGURE 2.2

(ii) If X has a negative binomial distribution as given by (Eq.2.2), then $E(X)=r/p$ and $V(X)=rq/p^2$.

(iii) If $Y_1....Y_r$ are independent random variables, each with the geometric distribution with parameter p, then $Y_1+...+Y_r$, the sum of these random variables, has the negative binomial distribution with parameters r and p. To see this, let Y_1 equal the waiting time for a first success, Y_2 the waiting time for a second success and so forth. Then $Y_1+...+Y_r$ is the total waiting time for the rth success and has the negative binomial distribution.

Remarks: (1) It is sensible that waiting time=n is less likely than waiting time=(n-1). In other words, it is less likely to have to wait longer for the first occurrence of the event. In order to have a waiting time of n, there must be (n-1) failures in a row, whereas for a waiting time of (n-1), there must be only (n-2) failures in a row, which is a more likely event. Thus it can be seen that P(X=n), the probability at trial n, decreases as n increases.

(2) The probability of success on or before n is the sum of the probabilities of the mutually exclusive events that the success occurred on trial 1 or 2 or 3, up to n. The probability of success by trial n, is the sum of the probabilities up to n, and this increases as n increases.

(3) These statements are consistent with the waiting time assumption that probability p of success at n, given no

success before n, is a constant, independent of time.

(c) The continuous case : the exponential distribution

It is sometimes mathematically simpler to idealize the probabilistic description of X, and to consider that we are dealing with a continuous random variable, which can assume all possible values in some interval $a<x<b$ (where $a=-\infty$ and $b=+\infty$). Suppose now that the observations are not restricted to occur within fixed intervals of time, but are made instead in continuous time. Let us suppose that there is an interval small enough that an event can occur, if it occurs at all, at most once. Then the waiting times assumption states that the probability of an event occurring in an interval from t to (t+h) is approximately a constant λh. This approximation improves as the size of h decreases, so that

$$\lim_{h\to 0} \Pr \frac{(\text{some event in } (t,t+h))}{h} = \lambda$$

and

$$\lim_{h\to 0} \Pr \frac{(\text{more than one event in } (t,t+h))}{h} = 0$$

and λ is independent of t, i.e. constant for any point t. λ is the probability of the occurrence of an event in an interval of time of small length.

A continuous random variable X, assuming all nonnegative values, is said to have an exponential distribution with parameter $\lambda>0$, if its probability density function is given by

$$f(t) = \lambda e^{-\lambda t} \quad \text{for } t>0 \quad =0 \text{ otherwise} \qquad \text{(Eq.2.3)}$$

Properties of the Exponential Distribution

(i) Graphical representation

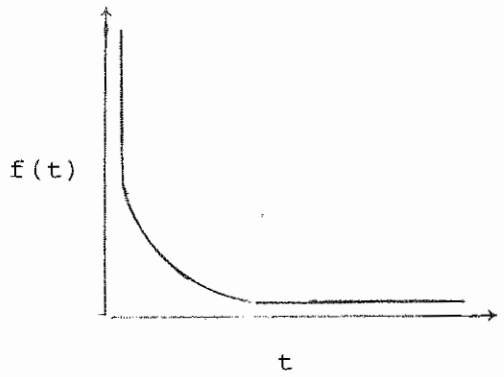


FIGURE 2.3

(ii) If X has an exponential distribution as given by (Eq.2.3), then $E(X)=1/\lambda$ and $E(V)=1/\lambda^2$.

(iii) The exponential distribution is the continuous analogue of the geometric distribution, and also has the property of 'no memory'.

(iv) The standard deviation is equal to the mean.

Remarks: Note again that (1) large waiting times are less probable than small, and (2) that the area under the

curve to point t, is the probability that the event will have occurred by time t. This probability increases as t increases. (3) All this is consistent with the fact that the conditional probability that the event of interest occurs for the first time in the interval of length h, given that it has not previously occurred, is a constant λh.

(d) The continuous case : the gamma distribution

To complete this exposition of waiting times theory, the gamma distribution remains to be described. This distribution is the continuous analogue of the negative binomial distribution, and is appropriate when dealing with multiple occurrences of an event. We want to know the probability of the occurrence of the kth event at precisely time t. We are interested in the probability that the sum of the waiting times for the first k occurrences of the particular event is equal to t.

Let X be a continuous random variable assuming only nonnegative values. X has a gamma probability distribution if its probability density function is

$$f(t) = \frac{\lambda}{(k-1)!} (\lambda t)^{k-1} e^{-\lambda t} \quad \text{for } t>0$$
$$= 0, \text{otherwise} \qquad \text{(Eq.2.4)}$$

This distribution depends on two parameters, k and λ, in which $k>0$ and $\lambda>0$. It is readily derived from (Eq.2.2) by

the method of moment-generating functions, a discussion of which is found in Mood (1950).

Properties of the Gamma Distribution

(i) Graphical representation

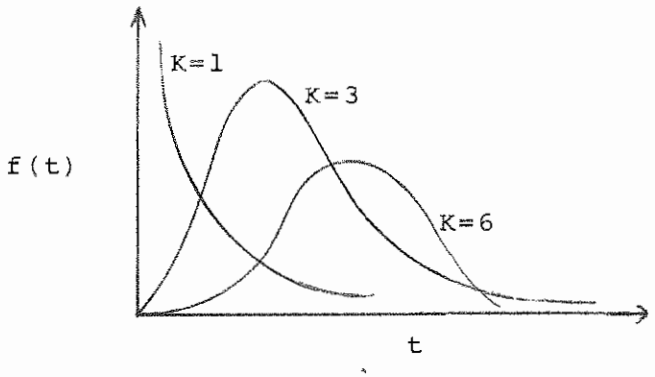


FIGURE 2.4

(ii) Note that when K=1, Eq. (2:4) reduces to $f(t)=\lambda e^{-\lambda t}$ (Eq.2.2). Thus, the exponential distribution is a special case of the gamma distribution. If k is a positive integer greater than 1, the gamma distribution is called the Erlang distribution.

(iii) If X has a gamma distribution given by Eq.(2:4), the expected value becomes

$$E(X)=k/\lambda \text{ and } V(X)=k/\lambda^2$$

(iv) The gamma distribution is the distribution of the sum[1] of k independent, identically distributed random

---

[1]The probability density function of the sum of k independent continues random variables is the convolution of the probability density function of the k random variables.

variables each having an exponential distribution with the same parameter λ. If the parameters of the various expenential distributions are different, the result is not a gamma distribution but some other distribution.

(v) When independent random variables are added together, the mean of the sum is the sum of the means, and the variance of the sum is the sum of the variances. With increasing k the gamma distribution becomes more nearly symmetrical, and the standard deviation becomes smaller relative to the mean. This property can be used to estimate the parameter k: since,

$$E(X)=k/\lambda=\mu, \text{ and } V(X)=k/\lambda^2=\sigma^2,$$

$$\mu^2=k^2/\lambda^2$$

and

$$\sigma^2=k/\lambda^2$$

so that $$k=\mu^2/\sigma^2$$

When the sample of observations is large, the obtained mean and variance, $\hat{m}$ and $\hat{s}^2$, may be used as estimates of the parameters so that

$$k=\hat{m}^2/\hat{s}^2$$

Alternative methods to this method of moments for estimating K will be discussed in Chapter 4.

(vi) The most general form of the gamma distribution arises when k is positive but not restricted to integral values.

2.2 Applications of the Waiting Times Model

Any experiment to which the waiting times assumption is pertinent, is a two-state Markov process with stationary transition probabilities. Markov processes form a very large and important class of stochastic processes with both discrete and continuous time parameters. A Markov process can be used to model any situation where the present state alone determines the future state, although the future state however is independent of the way in which the present state has developed out of the past. A stochastic process is said to be stationary if its probabilistic properties remain invariant when the process shifts to a new region of the time axis. A process that is not stationary displays change in its probability functions as time passes, and is classified as an evolving process. Nonstationary processes may or may not be Markovian in character.

The exponential distribution is the only continuous distribution characterized by the Markovian property of 'lack of memory'. Conversely, if a process exhibits no aging or decay, then the probability distribution of its

duration will be exponential (or geometric). For psychological purposes, it is often an inconvenience that Markov processes have no memory. However, Miller (1952) shows that the non-Markovian case, i.e. that the probability at trial (n+1)depends on the outcome of trials n and(n-1),can be made Markovian, since it is possible in principle to extend the Markovian definition to take into account as much of the past history of the system as is needed. On the other hand, Miller also points out that this is not always a useful procedure, and that most dynamic situations, as for example, those arising out of learning experiments, are better handled by other models.

Both the exponential and the gamma probability laws have been widely used in applied probability theory. The exponential distribution has been studied extensively because of its importance in physics and engineering. A number of interesting physical examples are discussed by Feller (1960). The behaviour of a variety of time-dependent processes has been able to meet the relatively strong requirements of this distribution. The exponential distribution was first used to describe the length of waiting times in telephone traffic situations. Since then, waiting time theory has been used to describe such phenomena as the life of an electron tube, the time interval between accidents,

studies of radioactive decay, flood flows and migration patterns.

The model has also been applied in the psychological literature. The simplest probability mechanism governing response latencies is one that assumes the response to have a constant probability of occurrence. In learning theory, the hypothesis of all-or-none learning states that learning is a unitary event, independent of past events, which has a constant probability of occurrence until it does occur (Estes, 1950; G.K. Bower, 1961). It may be that elaboration of all-or-none models to experimental data are required (Restle, 1962). Elementary processes may have all-or-none properties, but compounds of these processes cannot be unitary in nature. Instead, discrepancies between theory and data can be handled by considering that though the learning process is very complex, it can be adequately described if each stage of learning is thought of as an all-or-nothing process. Bush and Mosteller (1955) found that the solution latencies of rats learning to run down alleys could be described by the gamma distribution. Evidence from a range of experiments which included concept formation, paired associate and probability learning in children and T-maze learning in rats, strongly supported an all-or-none model which predicts a binomial distribution of response

with constant parameter p (Suppes and Ginsberg, 1963). In other words, the probability of a correct response remains constant over trials before the onset of learning, rather than showing a monotonic increase as would be predicted by an incremental learning model.

Other examples that appear to reflect the constant conditional probability mechanism can be found in several places in the psychological literature. The exponential distribution described inter-response times found at the onset of operant conditioning with certain types of reinforcement schedules (Mueller, 1950) but is not found during the later stages of conditioning (Anger, 1956). When subjects name the item that follows a test item on a short, recently memorized list, response time increases linearly with list length. The linearity and slope of the function imply that the test item is located in the memorized list by an internal self-terminating scanning process whose average rate is about four items per second (Sternberg, 1967). Records of spontaneous activity in a single spinal interneurone of a cat sometimes show an exponential distribution (Hunt and Kuno, 1959). Stephan and Mishler (1952) found that an exponential curve was a good description of the rate of participation in small face-to-face groups.

However, although these examples suggest that a

constant probability mechanism underlies response latencies, much of behavioural data in fact deviates from constant probability (see Felsinger, Gladstone, Yamaguchi and Hull, 1947). For example, McGill (1963) showed that reaction times to auditory stimuli depart systematically from linearity in the early stages of the distribution. The exponential curve was rejected in favour of the hyperbolic function as a more suitable description for the cumulative distribution of anagram solution times (Kaplan, Carvellas and Breinin, 1966). An excellent review of problems arising from the waiting times model is found in McGill (1963).

2.3 Application of The Model to Problem Solving

In Chapter 1, a brief discussion outlined a series of studies in which the waiting times model had been applied to the distribution of solution times for certain types of problems (Eureka), using both an individual and a group problem solving situation (J.H. Davis, 1961; Restle and Davis, 1962; Davis and Restle, 1963; J.H. Davis, 1964). The gamma distribution was used to model the distribution of times to solve multistage problems. A stage is loosely defined by the authors as a notion, idea, concept or part. An alternative interpretation would be to adopt information-processing terminology and consider a stage as a subgoal or

a collection of nodes. It is not clear from the reports referred to above whether the stages reflect the composition of the problem or the process. Perhaps it is fruitless to search for an answer to this question. Instead, it may be useful to consider the process as a behavioural mapping of the problem structure. We may view the structure of the problem-solving process as being determined by the structure of the problem that is being solved, although it remains independent and autonomous for investigation. Without a problem as stimulus, there would be no process, but the process can nevertheless be subjected to independent analysis. It makes sense to discuss these two aspects of problem solving separately: the problem in terms of its external and formal structure, the process in terms of the processing of 'chunks' of information that represent the different stages. Restle and Davis appear to confuse these two concepts. In one place, reference is made to the subject 'overcoming' various stages, and it is suggested that candidates for a stage may be "understanding the problem, hypothesizing etc" (1963,p.115). In a subsequent article, J.H. Davis discusses particular elements of a problem as potentially feasible stages, saying:

> Thus one might propose four distinct stages for this process if words rather than letters comprise stages.(1964, p.365).

Further ambiguity arises with the proposal that

> ...the number of stages to a wrong solution may be different from the number of stages to a correct solution. (1962, p.533)

It is unlikely that the same model can describe the processes to a correct and incorrect solution, although this would depend on the way in which the incorrect solution was reached. One subject might be unable to solve the problem as a result of executing a wrong decision at some branch of the problem 'tree', which he followed until some plausible, but incorrect, solution was attained. Another subject on the other hand, might be unable to solve the problem because he was working randomly in some irrelevant problem space that prevented him from grasping the true structure of the problem. Application of the stages model to non-solvers would be unproductive, since it would entail a restricted definition of non-solvers which it would be impossible to make operational.

This same argument could, on the other hand, be directed towards solvers. It might be claimed, for example, that there are some successful subjects who arrive at the solution by paths or stages quite different from those used by the majority of the group. This argument lacks plausibility: in the conventional problem-solving experiment, the structure of problems used by psychologists is usually

sufficiently defined that some degree of constraint is imposed on the behaviour of the subject. In addition to this, there is usually some systematic way of deciding when a certain solution is acceptable[1] (Minsky, 1963). This is not to imply however that the problem in question can be solved in only one way. We shall see in Chapter 6 to what extent individuals can differ in their choices at various decision-branches, and still solve the problem successfully. In the same chapter, we will determine to what extent the stages model can handle variability of behaviour.

In this thesis, the waiting times model will be confined to the task of describing the behaviour of successful problem solvers. This does not imply that it is therefore unnecessary to understand the processes that lead a subject astray. Through a detailed understanding of the processes and operations involved in solving a particular problem, it is possible to construct alternative paths, in order to point to the failure of certain mechanisms, or to

---

[1] A contrasting situation arises in experiments where creative problem solving is the object of investigation. Here, the goal or solution is not fixed, and interest centres rather on differences exhibited by individuals who have been given a common starting point. In such cases, it is plausible to view the subjects as working in idiosyncratic problem spaces which develop more freely in the absence of a restrictive problem structure. The stages model would clearly not be appropriate to describe such behaviour.

the erroneous application of others as an explanation of unsuccessful problem-solving behaviour. At an information-processing level, there exists a continuum between successful and unsuccessful problem solvers. In terms of the stages model however, the analysis will deal only with successful subjects.

Before examining the psychological meaning of the waiting times model and the assumptions it makes, it will be necessary to arrive at a working definition of a stage. The foregoing discussion has indicated the necessity for this definition to be two-fold.

Waiting times theory deals with the distribution of times, and this means that the model we have been describing is appropriate only for handling the process, and not the problem. It will be necessary therefore to distinguish between problem-stage and process-stage, so that 'stage' will be used to denote features of the problem-solving process, and 'module' will refer to features of the problem structure.

Definition

> A problem is said to be modular if it consists of one or more ordered modules, each representing a necessary sub-solution of the final solution. A module does not represent the smallest unit of analysis, since a given module may be composed of several elements. The ordering of the elements within the module may or may not be mutable.

A stage is the time to solution of a module when that time is exponentially distributed.

In an N-module problem, the total time to solution is the sum of the K independent times to solve each module, and is described by the gamma distribution.

The relationship between stage and module in the stage-module model is expressed in Figure 2.5.

The two parameters of the gamma distribution, K and λ, apply to the model in the following way:

(1) The stages model states that the distribution of solution times for an N-module problem is a random variable distributed gamma, and the sum of the K exponentially distributed random variables with identical parameter λ, which are the times taken to solve the N modules. Solving a complex problem can be represented as solving all of N modules successively in K stages.

(2) λ is a rate of probability, a probability per unit time. It is dimensionless, that is, the value of λ makes no simple sense, but depends on the units of time involved. λ is the conditional probability of the solution occurring in an interval of time of small length, given that it has not occurred previously. This means that the solution of a module is a unitary, all-or-none event, and that the conditional probability does not change, irrespective of how much time has been spent on the problem.

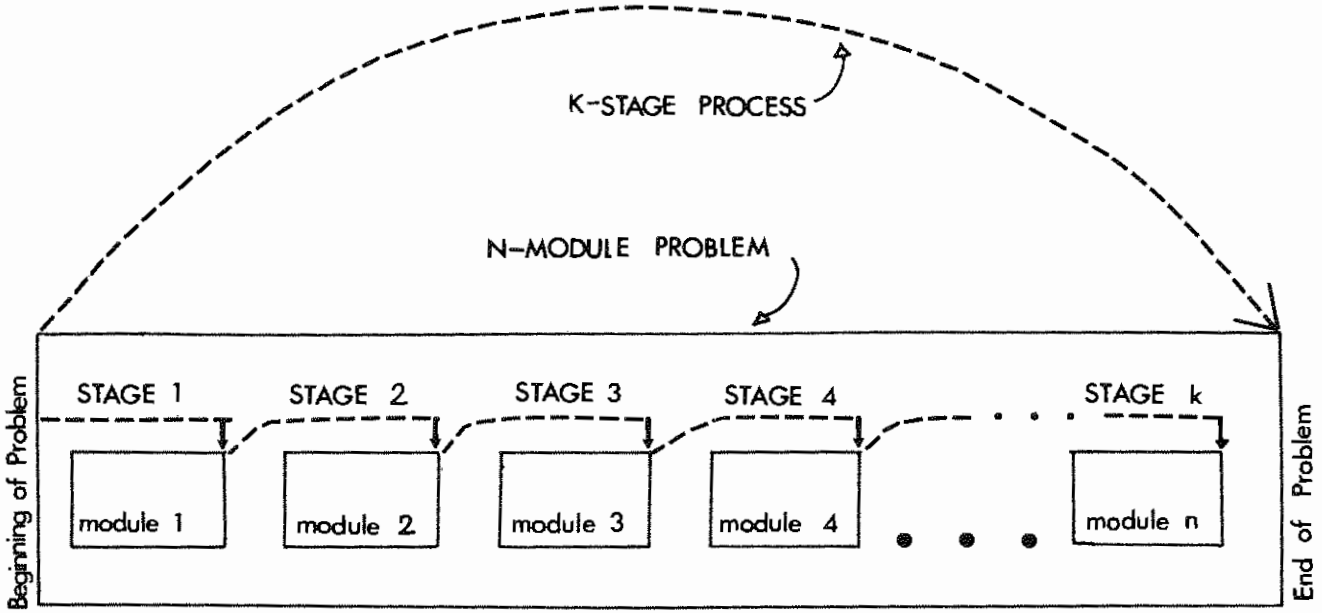


FIGURE 2.5: The stages-module model in which stages in the problem-solving process (K) are related to modules in the problem structure (N).

Assumptions:

When translated into a description of problem solving, the model as outlined above makes the following assumptions:

(1) The gamma distribution is the sum of K exponentially distributed random variables, identical with respect to parameter λ and mean. This implies that each of the K stages has the same conditional probability density, λ, for any probability interval in continuous time. In other words, each stage is identically distributed and each module is equally difficult, since they are each described by the same parameter value.

(2) The model assumes that λ is constant for all subjects. This implies that subjects are equal in their ability to solve the problem, since they each have probability λ of solving.

(3) The waiting times assumption of constant conditional probability within a module means that, given the module has not already been solved, there is a fixed probability that it will be solved in the next interval of time, irrespective of how much time has elapsed.

(4) The gamma distribution is the sum of K independent random variables. The assumption of independence means that the modules are assumed to be independently solved, so that the waiting time to the solution of one module bears no

relation to, and has no effect on, the waiting time to the solution of any other module. Work on Module 2 is assumed not to start until work on Module 1 has been completed.

To summarize, the assumption of a constant probability rate, λ, has three implications:

(i) for inter-module difficulty

(ii) for inter-subject ability

(iii) for intra-module difficulty.

In addition, there is

(iv) the assumption of independence of modules.

The model as it stands presents a picture of a group of problem solvers, each of whom resembles the other in his ability to solve a complex problem, composed of several equally difficult modules. In addition to this, the length of time spent on the problem (or module) has no bearing on the likelihood of solution within the next moment of time. Finally, the parts of the problem are related to the composite problem in such a way that success on one part of a problem does not affect the probability of success on any other part.

Clearly, many aspects of this depiction of problem solving will conflict with what is known either experimentally or intuitively about behaviour in general, or the solving of problems in particular. The severity of the model

is mitigated when it is remembered that any mathematical model is the embodiment of an idealized situation, and that some degree of flexibility is mandatory when applying it to behaviour. The remainder of this chapter will analyze the assumptions made by the model, so that the nature of the violations may be clarified.

The first assumption is the one least likely to present any major obstacles. Although it may be arduous in practice, it is quite possible in principle to find or construct a multimodule problem, with all modules equated for difficulty. J.R. Hayes (1966) developed 'spy problems' which were both homogeneous, since all steps were equally difficult, and modular, since it was possible to build them up or down for variations in complexity. These problems basically involved the learning of paired associations between words, and it would be more complicated to meet the criteria of homogeneity and modularity in devising a complex mathematical task. Since it is theoretically possible to control inter-module difficulty, the stages-module model can be investigated under systematic violations of this assumption (see Chapter 5). Secondly, it is entirely untenable that individuals are identical in their problem-solving ability, and differences will certainly increase with the use of increasingly complex problems. Any model of

problem-solving must explicitly account for individual differences,[2] unless it can be shown that this assumption may be violated within certain limits without gross distortions to the fit of the descriptive distribution.

The third assumption states that the time spent on a problem, or on a module unit of problem, has no influence on the probability of solving it in the next moment of time. This waiting times assumption gains in plausibility when it is augmented by the remark that although the probability of solution in any interval remains constant, the probability of solving a problem by a certain time increases as that time increases. (See Remarks on pp.46-47). The tenability of the assumption of a constant conditional probability mechanism has already been discussed in terms of evidence arising from

---

[2] Bush and Mosteller (1955, pp.319-321) do not assume either the same starting point, nor the same probability rate for each animal in their analysis of data obtained by Weinstock (1954) in a runway experiment. These two parameters are estimated separately for each rat, although the parameter k is estimated for the whole group. By contrast, on the assumption that all subjects have the same conditioning parameter, c, G.K. Bower (1961) derived the following distribution for the trial n' on which the last error occurs:

$$\Pr(n'=k) = \begin{cases} b/N & \text{for } k=0 \\ b(1-1/N)(1-c)^{k-1} & \text{for } k>1 \end{cases}$$

where $b = \dfrac{c}{1-(1-c)/N}$

The goodness-of-fit test for this distribution of last errors is a test of the null hypothesis that subjects are homogeneous with respect to their conditioning propensity.

previous research in Section 2.2. It may be that the concept of a constant probability rate is misleading, and that a model with fluctuating or different time constants is more appropriate. In an investigation into inter-response time between successive verbal associations (Bousfield and Sedgwick, 1944), it was found that there was a step by step drop in the time constants. If the subject is viewed as searching randomly through a storage consisting of a constant number of alternatives according to a sampling with replacement model, then the time constant for any response is proportional to the number of unused alternatives at that point. The time for examining each alternative is exponentially distributed, and the process runs on until an unused alternative is discovered (McGill, 1963). If, on the other hand, the model describing the selection of alternatives from storage is closer to that of sampling without replacement, then it would be natural to predict uniform time constants. The findings reported by Bousfield and Sedgwick are supported by Hunter (1959) who noted this same temporal phenomenon in the solution of anagrams, so that the longer a subject spent on the problem without solving it, the longer he would have to wait before he did solve it.

An alternative description of problem solving might be provided by the pure birth (Yule-Furry) process

(Bharucha-Reid, 1960,pp.77-78) which is a latency process that speeds up i.e., the time constants increase, with time. By contrast, a pure death process is characterized by time constants which decrease in steps of size λ with each response, so that the mechanism appears to be slowing down, as in the Bousfield and Sedgwick study. Perhaps a combination of these two processes might provide the most adequate model:[3] summaries of well known combination chains of birth and death components are found in Bharucha-Reid (1960) and Feller (1960). McGill (1963) has developed the generalized form of the gamma distribution where t is summed over the k+1 stages, each exponentially distributed with the time constants arbitrary and all different. The density function for this distribution, which is the weighted sum of the exponential components, is given by

$$f_k(t) = \sum_{i=0}^{i=k} C_{ik}\ \lambda_i e^{-\lambda_i t} \qquad \text{for } t \geq 0 \qquad \text{(Eq.4.5)}$$

and the weights are computed from the time constants, where λ_i is the time constant of the ith event in the chain

---

[3] The values presented in Chapter 4 (Table 4.10 and 4.11) suggest that increasing time constants may be appropriate for describing the beginning and end of a problem, and decreasing time constants more suitable for the middle of a problem.

$$C_{iK} = 1 \Big/ \prod_{\substack{j=0 \\ j=1}}^{j=k} (1 - \frac{\lambda i}{\lambda j}) \qquad \text{(Eq.4.6)}$$

The fourth and final assumption, that of independence of the random variables is violated in any problem where the modules are sequentially arranged so that Module 2 must be solved after Module 1 because the information contained in the first module is a prerequisite for solving the second. The model assumes that when a subject is working on a given module, he cannot be working on any other module at the same time. This assumption is probably frequently violated, since a subject will abandon work on one part of the problem, if progress seems slow, in order to tackle another part. In this sense, a stage is usually 'impure', in that its time contains traces from other stages. It is also assumed that information cannot be carried from one module to another, either in the form of specific information, or in the form of generally applicable hypotheses and heuristics that have proved effective on earlier modules.

Both assumptions (2) and (4) are violated in varying degrees by the problems used in this study. It will be shown (Chapter 4) that not only do different modules have different probability rates associated with them, but the higher rate in one module is, in some cases, directly

determined by the fact that a previous module has been solved. This occurrence constitutes a violation of the assumption of independence as well as of a constant λ.

Restle and Davis (1962, pp.530-532) point out that the stages model does not apply to the situation where Module 1 or 2 or 3 have to be solved. The modules must be solved sequentially. If the order is variable, then a new distribution arises. The authors substitute 'parts' for 'stages' (our modules) in the case where the problem is such that "ideas for different parts can arise freely and independently". It is not clear what is meant by this, but presumably parts differ from modules in being internally less structured and temporally distinct, as well as in lacking the constraint of sequential ordering. The solution time for any part will have an exponential distribution,but until the first part has been solved, k parts remain to be solved. This means that the rate until the first part is solved is $k\lambda$, for the next part $(k-1)\lambda$, until, for the last part the rate is λ. Since the mean of each waiting time is independent, the total mean is the sum of k waiting times with means $1/\lambda$, $1/2\lambda$, $1/3\lambda$,....,$1/k\lambda$. This gives us that

$$E(X)=1/\lambda,(1+1/2+1/3+...+1/k).$$

Testing the model

The descriptive power of the model is tested in three ways:

(1) The number of stages in the total waiting time is determined using moments or maximum likelihood methods of estimation (Chapter 4).

(2) This in turn is a direct indication of the number of modules in the problem which is independently verified from the protocols of individual subjects.

(3) Appropriate gamma distributions are fitted to the total solution times; appropriate exponential distributions are fitted to individual stage times.

There are aspects of the model which are either impossible, impractical or forbiddingly expensive to test by experiment. It may be known which assumptions are being violated, but the effects of this on the model are impossible to substantiate with evidence available only from one set of experiments. We want to ascertain whether the modules are independent, and this can be tested by measures of correlation. It is likely that the size of λ is variable both for subjects and for stages, but conventional techniques of analysis will not provide the information about the size and direction of distortion, if any, that these violations produce in the descriptive fit of the model.

It is also impossible to quantify the extent to which λ varies between subjects and stages. These questions may however be answered through the application of Monte Carlo simulation methods, and Chapter 5 will be devoted to the results of testing the stages-module model by these techniques.

CHAPTER 3

PLAN OF THE STUDY

It is the aim of every experimenter to design a study that will maximize the quantity and quality of information necessary to examine his hypothesis. If his intention is to investigate the processes by which problems are solved, then he must select a set of potential problem solvers, and a task or series of tasks that will be appropriate to that group.

It seems reasonable to restrict the choice of subjects to those who seem most likely, a priori, to be able to solve the problems in which one is interested. It is inefficient to build a model of problem-solving behaviour based on observations of poor problem solvers who have a low probability of success on particular problems.[1] It would be equally unrewarding, on the other hand, to chose problems that are certain to be solved by any randomly selected individual.

Which comes first then, the problem or the subject? A valid procedure would be to find a problem that is

---

[1] Note however, the profitable use of insoluble problems by Paige and Simon (1966).

sufficiently complex to produce behaviour rich in analytic possibilities, sufficiently flexible for the display of individual styles of attack, and with sub-parts sufficiently ordered that the solution is not intuitively obvious from the beginning. If the problem is too difficult, it will be impossible to find enough successful subjects for a legitimate study. If it is too easy, the subjects will solve the problem too quickly. Since this study employs the technique of verbalization to 'externalize' thought, an easily solved problem might result in thinking that proceeded too fast to be tagged at each step by a speech monitor. Instead, the subject might be forced to reconstruct his thoughts only after the problem was solved, or to slow down his thinking to a speed at which verbalization could function as a parallel process. This would produce a distorted and artificial representation of the internal cognitive situation. At an optimal level of problem difficulty, the verbalization could be expected to synchronize with the mental activity, with the process of verbalization becoming a reflection of, rather than an influence on, the process of thinking. This suggests that not all problems would be suitable for use in verbalization experiments.

An additional proviso for an appropriate class of problem has to be made in the present case: the problem

should have an intrinsic structure to which the stages model could reasonably apply. Many problems are not organized in such a way as to permit a breakdown into an infra-structure consisting of parts or stages or some other kind of smaller unit. At the same time, the class of problem used must not be so esoteric as to make it impossible to generalize to other classes of problems. Finally, if it is feasible,there is much to be gained by extending research on problems and materials used by other workers.

Assuming that a suitable problem has been found, it remains to select an appropriate solver. If we allow ourselves to sample outside some defined universe or problem solvers, we will not learn very much about problem solving. If instead, we sample only from among those who are certain to be able to solve the problem, it is likely that the result would be a homogeneous group with little or no variance.

It is hoped that this study has succeeded in selecting a class of problem that proved challenging, but nevertheless soluble, for a majority of subjects sophisticated enough to be able to verbalize their thinking without undue difficulty.

This chapter begins with a description of the problems used in the study (3.1), and is followed (3.2) by an

account of the subjects involved in each of the experiments (I,II and III). The next section (3.3) outlines the procedural details for each experiment, while the last section (3.4) briefly summarizes the data gathered in the study, analysis of which provided the results discussed in subsequent chapters.

3.1 The Problems

General requirements for problem selection have been reviewed in the preceding section; a specific criterion for this research was that no specialized training should be required of the subjects. More particularly, no mathematical knowledge, beyond the rules of simple arithmetic,would be necessary to solve the problems. Since the study was not designed to investigate mathematical problems *per se*, it was important to eliminate spurious differences that might originate from variations in amount of mathematical training. With this restriction, individual differences would be revealed through the ways and skills with which basic tools were applied to situations.

It would have been fruitful to use the problems given by Restle and Davis (1962) to their subjects. Two of their problems were rejected since they lacked the required degree

of complexity.[2] The third problem, a variant of Luchin's Water Jar problem (Luchins, 1942) was tested in a pilot study, but even this more difficult problem did not prove sufficiently demanding to produce protocols of any substance.

Rather than add another species of task to the enormous variety already used in psychological experiments, it was decided to select examples of cryptoarithmetic problems,[3] a type of problem used by previous investigators (Bartlett, 1958; Newell, 1966a), and which satisfied the specific and general criteria of problem selection.

In cryptoarithmetic problems, the solution is reached by decoding an arithmetical task expressed in alphabetic form, into its numerical equivalent. An example of an addition problem is

```
 SEND
 MORE
------
MONEY
```

Specific rules can vary, but usually (and always in these experiments), each letter stands for a digit, and each digit is represented by only one letter. Thus a one-to-one correspondence obtains between letters and digits. There

---

[2] The 'Rope' problem is very simple, requiring for its solution the attainment of only a single idea. The 'Word Tangle' problem requires a 'yes-no' answer, followed by 'Why?'. Although this could be reworded, the basic problem still seemed to lack sufficient challenge.

[3] Sometimes known as alphanumerics or cryptograms.

may or may not be a unique solution, and the problem may contain up to ten different letters. Solutions do not depend on any non-numerical rules, as, for example, one that would link letters to numbers in some alphabetic sequence, so that A=1, B=2 etc. The configurations of letters are decoded entirely by the application of basic arithmetical rules.

The problems were obtained from the files of a local newspaper which published these and other puzzles; none of the subjects had previously seen any of the problems used in the study. A battery of three problems was considered sufficient, since it was envisaged that some subjects might work on them for a considerable period before finding a solution.

Problem 1

DONALD
GERALD
ROBERT

This problem was used by both Bartlett (1958) and Newell (1966a). Subjects were told that D=5. With this constraint, the problem has a unique solution.

Problem 2

SIR
× I
PEEP

This multiplication problem has four known solutions.

Problem 3

```
  HANNAH
    MUST
   STUDY
 TUESDAY
```

This addition problem has only one solution.

Both the addition problems contain ten different letters, whereas the multiplication problem has five. The third problem was considered by most subjects to be much harder than the other two. This fulfilled the experimenter's intention of maximizing the range of difficulty presented by the problems, in order to obtain as large a sample of behaviour as was possible in a limited study.

3.2 The Subjects

The design of the study required that some subjects solve the problem verbally in order to provide data for deeper analysis of the stages model, and that other subjects solve the problems nonverbally, to allow the experimenter to investigate the effects of verbalization on solution times. This meant that a large number of suitable subjects would have to be recruited.

Because the problems required particular expertise (though not formal training), subjects could not be selected at random, but would have to be chosen from some specialized population, whose dimensions could not be known

precisely in advance. Although we could be confident that those who eventually solved the problem were from this population, it was not possible to ensure that everyone in the population had a chance of entering the experiment. In this sense, the set of people in the experiment do not constitute a sample, and the results cannot be generalized to any population of problem solvers. Since all the subjects in the study were volunteers, the only way in which 'selection' occurred was in the sense of self-selection by the volunteers. At the final level, the groups consisted of only those subjects who had successfully solved a particular problem.

The only consistent criterion applied to the experiment was that the numbers in the verbal and nonverbal groups should be as large as possible, and as nearly equal in size as feasible. The constraints of the experiment therefore led to uncontrolled variation in the characteristics and backgrounds of subjects. Restle and Davis (1962) do not account for individual differences in their work, but other research in problem solving, which does not employ the stages model, treats these differences in some detail (Duncan, 1959, pp.412-417). Under ideal conditions, it would have been desirable to combine what is known about individual differences with the stages model, and observe

the various interactions of these factors. In Chapter 4, some suggestive data will be presented on individual differences. Although these were in the expected directions, this thesis cannot discuss these interactions with any confidence.

In summary, although we are aware of the importance of individual differences, and their potential influence on the outcome of the experiment, the various groups of subjects were combined within each problem with regard only to the verbal-nonverbal distinction.

Subjects were drawn from various places, and were tested over a period of several months. In the first stage of the experiment (Experiment I), letters were sent to members of the Mathematics and Statistics Departments of the University, asking for volunteers. Another group (Experiment II), was recruited with the co-operation of the headmasters of various secondary schools in Canberra, who made available their highest level Mathematics students in the Fifth Grade, during school hours. A final group (Experiment III) was collected by asking for volunteers among academics and computer programmers, as well as recently graduated Sixth Grade students, who had had some degree of mathematical training in their final year. The experimental conditions varied from group to group, and each is discussed below in detail.

a. EXPERIMENT I:

Adults (individual condition): verbal and nonverbal

Subjects were alternatively assigned to a verbal or nonverbal group according to the order in which they contacted the experimenter. The subjects themselves had only a slight idea about the content of the experiment, since they had been told that some degree of mathematical or arithmetical facility was necessary, and that the experiment could last up to 2½ hours. They were not told to which experimental group they had been assigned.

Out of a total of 41 who volunteered, 36 subjects were successful on at least one problem. Of the five who were rejected, four were unable to solve any of the problems, and one, although successful, could not be used, due to a fault in the recording equipment during his session. The 36 subjects formed two groups of 18 each. One group was tested under instructions to verbalize (verbal) and the other under normal conditions (nonverbal).

The subjects who volunteered were all males. They included university undergraduate and graduate students, faculty members, computer programmers and school teachers. About half of them had previously solved cryptoarithmetic problems.

b. EXPERIMENT II

Students: (group condition): nonverbal

This second experiment was performed in order to enlarge the size of the groups of Experiment I which proved too small for any reliable statistical testing: one of the problems was solved by only 13 subjects in each experimental condition. The multiplication problem was not used in Experiment II and III, since the experimental sessions tended to last too long. It was decided to shorten the time involved by excluding one of the problems. Since the two addition problems could be compared profitably on the same dimension of complexity, whereas the multiplication problem appeared to lie outside this dimension, and to be less amenable to stage-module treatment, the latter problem was eliminated.

This nonverbal group consisted of high school students who had just completed Fifth Grade. They were both male and female, aged 16-18 years. Only those students who had performed at the two highest mathematics levels (1 and 2F/2S) were used. Due to school regulations, these students had to be tested in the classroom during school hours, which meant that experimental sessions were limited to seventy-five consecutive minutes. It also restricted the subjects to being in the nonverbal group, since it was not

possible to transport the recording apparatus to the schools. With a few exceptions, this meant that the students only had time to work on one of the problems. Although the experiments were performed during school hours, the annual examinations were over, and those who were unwilling to participate would have been excused. The motivation among all subjects was very high, and many asked for another problem to take home; it is unlikely, therefore, that any importance can be attached to the fact that the subjects in this group were not, strictly speaking, volunteers.

The students were tested in six high schools. The classes consisted of groups ranging in size from 9 to 22. A total of 106 subjects were tested, of which 56 were successful on the problem they had been given. Six of the 56 were able to complete a second problem successfully. This brought the number who solved Problem 1 (DONALD) to 37 and Problem 3 (HANNAH) to 25.

c. EXPERIMENT III

Students and adults: (group condition): verbal

Since the high schools appeared to be a rich source of otherwise very scarce subjects, the registers for Sixth Grade students who had just completed their final year of school were used to select those who had worked at any level of mathematics. A letter was sent to 194 students, asking

them to volunteer. They were offered $2.50 for spending two and one-half hours as subjects. It is possible, but unlikely, that this reward created uncontrolled motivational differences between this group and any of the others to the extent of influencing a subject's ability to solve cryptoarithmetic problems. Of the 41 initial respondents to the recruiting letter, 27 actually acted as subjects in the experiment. Of the 27, 18 were successful on one problem, and 9 of the 18 completed two problems satisfactorily. These student subjects were male and female, aged 17-19 years.

To augment the size of this group, eight computer programmers and mathematics graduate students volunteers (adults) were included in Experiment III, and paid the same amount. Seven of these were successful on at least one problem, five on both problems.

The structure of the groups in the experiment, along with the success rates among subjects who were tested, is presented in Table 3.1. These rates may be compared to those of J.H. Davis (1964) among undergraduates in an introductory psychology class, where 78% solved the relatively easy 'Murples' problem, and 57% were successful on the 'Card' problem. The rates in the present study, given more difficult problems, and both more and less sophisticated

subjects, appear to be congruent with Davis' findings.

TABLE 3.1

Number of subjects tested in each experiment and who successfully solved each problem with resulting success rates for each problem

Experiment	Number of Subjects Tested	Number of Subjects Successful on Each Problem:			Success Rates (%) on Each Problem:		
		1	2	3	1	2	3
I	41	34	34	26	83	83	64
II	106[4]	37/66	-	25/61	56	-	41
III	35	23	-	16	66	-	46
		N=94	N=34	N=67			

3.3 Procedure

(a) EXPERIMENT I

The first experimental sessions were held from December 1968 to February 1969. Each subject had contacted the experimenter to volunteer, and to make an appointment for the sessions which were usually held in the evenings.

---

[4] 106 subjects included 66 who were tested on Problem 1 and 61 who were tested on Problem 3. This total is 127, but 21/127 were tested on both problems, and 6/21 succeeded on both.

The subject was seated in a small cubicle. A microphone, attached to a cord, was placed around the subject's neck. A tape recorder stood on a table directly outside the cubicle, along with an oscillator connected to the second input of the tape recorder. This emitted signals lasting 0.5 seconds, every 20 and 40 seconds, and a signal of 1.0 second duration every 60 seconds. The switch was driven by a synchronized motor. The signals were not audible while a recording was being made. They were intended to be heard only while the tape was being transcribed, so that the protocol could be sectioned every 20 seconds. In front of the subject were some scratch paper and pencils and the problem sheet, face down, on which were written the rules for the solution and the problem itself (Appendix A).

During the instruction period, the experimenter played a tape recording of approximately one minute duration, which was a segment of a tape made by a subject verbalizing during the pilot study. The segment was carefully chosen so as not to transmit any information which might be useful in solving the problems, even though the subject in the pilot study had been working on a problem not used in the main study. The intention in playing the tape was to minimize any anxiety or foolishness the subject might feel at the prospect of 'thinking aloud' and being recorded.

When he was ready to begin, the subject was told

"I am going to give you a series of three problems, one at a time, and I would like you to think out aloud into the microphone as you work them out. The problems involve decoding alphabetic problems to find their correct numerical solution. You may already be familiar with this kind of problem. Here is an example. SHE
+BET
CASH was shown. You are asked to assign digits from zero to nine to the letters in the problem so that each letter corresponds to a unique digit, and different letters represent different digits. You may also assume that zero cannot be assigned to the leading left-hand digit in any row. These rules will be repeated on the problem sheets I have given you. Some problems have more than one solution, but don't worry about getting more than one. Even though I shall be timing you, there is no time limit, and your answer will be useful to me even if it is incorrect. Have you any questions so far?
I am interested in the way you set about solving the problems rather than in whether you can solve them successfully or not. I want you therefore to feel free to talk as much as you possibly can. At first it will probably seem a little strange to you to have to do your thinking out aloud, but I am sure you will get used to it in a short time. I shall now play you a short tape of somebody else working on this problem we have in front of us. You will not be given this particular one, but I just want you to get an idea of the sort of thing I want you to do.(Play tape.) Do you have any questions? Here is the first problem. It is an addition one. Please dont be discouraged if you feel you are not getting anywhere because this is not an easy problem. Turn over the paper when I tell you to. You will hear me start the tape recorder at the same time. When you are satisfied that you have completed the solution, please say 'I have finished'. Remember again that I am interested in anything and everything you are thinking about as you work on the problems. Please turn over the page now."

The experimenter closed the door, started the tape recorder, told the subject to begin and simultaneously started a stopwatch. The subject was monitored through earphones from time to time to ensure that a satisfactory level of recording was maintained. He was not interrupted or reminded to verbalize during the experiment, since it was felt that this would disrupt his thinking, and at the same time inform him that he was being listened to while he talked, which could well prove to be inhibiting. To interrupt would also create problems later when parts of the protocol were being timed, since the time taken by the experimenter speaking would have to be deducted from the total time. If a subject had seemed reluctant to verbalize, or if his rate of verbalization seemed to be low, he would again be encouraged, during the interval between problems, to talk as much as he could. This often remedied the situation.

When the subject announced that he had finished,the experimenter stopped the stopwatch and tape recorder and went in to check the solution. The subject was told whether it was right or wrong. The problem sheet and all scratch work were kept in case ambiguities arose during analysis. After a few minutes rest, he was given the second problem and told that this was a multiplication problem with more

than one solution. When the third problem was presented, he was told that D=5. All subjects were presented with the problems in the same order:

(1) HANNAH
(2) SIR
(3) DONALD

If the subject took more than sixty minutes, he was stopped, and it was suggested that he might like to try another problem instead. He was again reminded that the recording was useful even without a correct solution. These reassurances may seem elaborate, but they were made because the problems were objectively difficult, and many of the subjects gave a distinct impression of being threatened by the possibility of failure. This was probably heightened by the fact that they had volunteered on a self-judged basis of mathematical ability.

There were three ways in which a subject could fail to be included in the results: (a) by producing an incorrect solution (b) by spending more than sixty minutes on a problem, and (c) by abandoning the attempt to solve the problem. When the subject had completed, or attempted to complete, the three problems, he was asked whether he had previously tried to solve cryptoarithmetic problems.

The procedure for the nonverbal group was substantially the same as for the verbal, except that all reference

to verbalizing was eliminated from the instructions, and the sample tape recording was not played. The subject was timed in the same way, i.e. a stopwatch was started simultaneously with instructions to turn over the problem sheet. The subject was told that he would be asked some questions about his procedure after he had finished, and that his answers would be recorded. He was asked to say 'I have finished' in a voice loud enough to be heard by the experimenter. The stopwatch was stopped as soon as he said this. Whether he had solved the problem or not, the subject was then asked to reconstruct, in as much detail as possible, the way he had solved, or tried to solve the problem. This was recorded on tape. The experimenter asked questions when necessary, in order to maximize the amount of information for each subject. The subject was asked to remember any errors or false tracks he had made, but the experimenter placed particular emphasis on the order in which the subject had correctly assigned digits to the letters. When he was unable to add any more to his description, he was presented with the next problem. This procedure was repeated, and the third problem presented. The same criteria for failure were used as for the verbal group.

(b) EXPERIMENT II

This experiment was conducted in the classroom during school hours. Since only 1¼ hours was available for testing, there was rarely time for more than one problem to be presented. The first groups were given Problem 1 (DONALD), and when a sufficient number had solved it successfully ($N \geqslant 30$), the remaining classes were given Problem 3 (HANNAH). No other criterion was used for deciding which subjects would be presented with which problem. Since all the high schools were State schools following the same syllabus, and since all the pupils were performing at approximately the same level of mathematics, it was assumed that this was a homogeneous group with respect to problem-solving ability.

It was not possible to replicate the individual conditions of Experiment I. However, the subjects appeared absorbed in the tasks, and as there was no communication between them, it is difficult to see what interference could have been caused by this group testing situation. Ideally of course, the same situation should have been used throughout the study, but this was not possible. Subsequent research may benefit from the experimenter's experience that individual testing is a very time consuming procedure, especially when demanding problems are used.

Group size varied from nine to twenty-two, according to how many students arrived for the session. The instructions given to the subjects were substantially the same as those given to the nonverbal group in Experiment I, and the same personal information, including the subject's age, was collected. Anticipating the possibility of confusion if the experimenter was responsible for timing each individual in a group, in this experiment the subject timed himself. Instructions were given to start the stopwatch simultaneously with turning over the problem sheet, and for the watch to be stopped as soon as the subject was satisfied that a solution had been attained. These instructions were repeated, to ensure that no subject would forget to stop his stopwatch at the appropriate time, thereby becoming invalid for inclusion in the analysis. Subjects were told that they would be asked to reconstruct and write out their solution paths on the problem sheets after the problem had been completed. It was emphasized that this should take place only after the watch had been stopped.

The subject indicated that he had finished the problem by raising his hand. The experimenter checked the solution, and recorded the time on the stopwatch. The subject was not permitted to renew his efforts if the solution was incorrect, but if sufficient time remained, he was

presented with the next problem. It was usually necessary for the subject to be reminded to write out his solution steps, which he frequently said he was unable to recall. He was urged to try to reconstruct at least the order in which he had replaced letters by digits. Again, the subjects seemed to experience difficulty in complying with even these modified instructions. If the subject was willing, and if sufficient time remained, he was presented with the second problem.

(c) EXPERIMENT III

Subjects were tested in a language laboratory, which consisted of a large room, divided into 53 soundproofed booths, with partitions three feet high on either side,and in front of the subject. He was therefore relatively insulated from other people in the room. Each booth was fitted with recording equipment, and the subjects wore earphones with an attached microphone. The earphones and the soundproofing minimized any outside interference, although the subjects were aware that they were being tested in a group.

Three evening sessions were held, with ten to fifteen subjects in each group. The majority of the subjects were recent high school graduates, male and female, but the last session consisted mainly of graduate students and computer

programmers, who were, with one exception, male.

The procedure and instructions for this experiment followed that for the verbal group in Experiment I, except that subjects were required to time themselves. Each subject was given a stopwatch which he was asked to start as soon as he turned over the problem sheet. A technician was present to instruct the subjects in the use of the tape recording equipment, and it was emphasized that the machine should be running before the stopwatch was started. One subject omitted to do this, and no recording is available for her (see Table 4.10). It was possible to monitor any subject from the master console, but, as in Experiment I, the experimenter did not interfere with the ongoing process in any way. To gain extra verbal output by such means was considered to be less valuable than to allow individual differences in verbalizing styles to emerge unimpeded. This also ensured that the stages model could be applied without ambiguity.

Each subject wrote on the problem sheet his name, age and whether he had previous experience with cryptoarithmetic problems. To eliminate possible confusion in the identification of the tapes, each subject labelled the tape spool with his name, and spoke his name at the beginning of the recording, before starting the stopwatch. He was instructed

to say 'I have finished' when he completed the problem, and then to immediately stop the stopwatch and the tape recorder. He then raised his hand, and the experimenter presented him with the second problem sheet which he turned over when he was ready. The procedure was repeated. Problem 1 was presented before Problem 3 in Experiment II.

Some subjects complained of discomfort arising from the earphones, but only two or three in the entire experimental series mentioned any difficulty in verbalizing. This will be discussed in Chapter 7.

The experimenter conducted Experiment I alone, but, due to the problems of administering the group sessions, an assistant was used in Experiments II and III.

3.4 The Data

The data gathered in the experiments consisted of (1) solution times for verbal and nonverbal subjects (2) transcriptions of the verbal behaviour of verbal subjects (3) post hoc reconstructions of nonverbal subjects and (4) auxiliary personal information.

The tapes produced by subjects in the verbal groups of Experiment I and III, were fully transcribed with periods of silence noted. The signalling device used in Experiment I made it possible to section the protocols in units of

twenty seconds each. In Experiment III, it was decided to register intervals of one minute instead, since the measurements used in Experiment I proved to be unnecessarily fine. It was not possible to incorporate an oscillator into the equipment used in Experiment III; consequently, the transcriptions were segmented with the aid of a stopwatch. (This timing procedure was essential for a rigorous testing of the stages model, and is discussed in Chapter 4.) Since it was not economical to note the time at which each word was said, it was assumed that words between any two minute-markers were evenly spaced, and the time at which a particular word or phrase occurred could thus be estimated from its relative position within a one-minute period.

Post hoc reconstructions provided information about solution paths and letter-digit substitutions for the nonverbal groups; this was contained in the problem sheets for Experiment II, and in the tape recordings, made after each problem was solved, for Experiment I.

The next chapter discusses the results of analyzing this data.

CHAPTER 4

QUANTITIVE RESULTS OF THE STUDY

The fabric of behaviour reveals a spectrum of qualities according to the lens through which it is viewed. At one level, the problem solver appears in his conventional role as a contributor to the variance of the total distribution of times taken by a group of individuals solving a problem. His significance is diminished when,at a second level, interest is focussed on the number of modules characterizing a particular problem, or when estimates are made of the number of component processes comprising a total solution process. At the third, and finest level, there is a palpable increase in the saliency of the individual subject when he is seen as a processor of information, generating highly complex cognitive structures as he searches for the solution to a problem.

In this chapter, the results of the experiments will be analyzed at the first two of the levels described. Chapter 2 provided a theoretical framework for the definition of both stage and module. The criteria developed there will be applied to assess the viability and limitations of the concepts of stage and module when these are couched in purely statistical and numerical terms. This is considered

an intermediate level of analysis. The quantitative content of this chapter, in particular the $\hat{k}$ estimates provides a base from which to embark onto the third and deepest level of investigation of problem-solving behaviour (Chapter 7). Chapter 6 contains the results of this intermediate level of analysis, and will describe the structural underpinnings of the estimates calculated for number of stages and, particularly, for number of modules.

The first section of the present chapter (4.1) presents the results of the study in terms of solution times and variance, with emphasis on the differences found between various groups of subjects, when they are distinguished by experimental conditions, age and previous experience with cryptoarithmetic problems. Section 4.2 presents the K estimates, using two methods of estimation. Section 4.3 contains a series of cumulative relative frequency distributions of solution times required to solve (a) the problem as a whole and (b) individual modules, fitted by the appropriate theoretical distributions. Section 4.4 discusses the adequacy of the stages-module model when it is applied to data obtained from protocol analysis.

Solution times provided the raw data for this aspect of the research. The subjects from Experiments I, II and III were treated as a single group from which variables of interest were selected for analysis.

4.1 Solution Times

(a) Experimental conditions: verbal and nonverbal

Means and variances were computed separately for the verbal and nonverbal groups, and pooled estimates were obtained by combining all subjects who had solved a given problem. In Table 4.1, where these results are reported, we find striking similarities between the solution times of the subgroups, and erratic, although never significant, differences between subgroup variances. This suggests that verbalization does not affect the mean solution time of a relatively large group of problem solvers, but may have a limited bearing on the performance of an individual.

TABLE 4.1

Mean solution times (minutes) and variance for three experimental problems for verbal, nonverbal and pooled verbal-nonverbal groups

Problem	Condition	N	Mean Time	t	Variance	F
1 (DONALD)	Verbal	41	26.30	0.21	272.26	1.47
	Nonverbal	53	26.98		185.62	
	Verbal and Nonverbal	94	26.68		221.00	
2 (SIR)	Verbal	18	20.62	0.22	65.12	1.86
	Nonverbal	16	19.88		121.10	
	Verbal and Nonverbal	34	20.27		88.73	
3 (HANNAH)	Verbal	29	34.38	0.61	269.76	1.00
	Nonverbal	38	36.81		269.52	
	Verbal and Nonverbal	67	35.46		259.15	

If mean solution time (using pooled data) is taken as an index of problem difficulty, we observe that variance increases as the problem becomes more difficult, so that Problem 2 (SIR), with a mean solution time of 20.27 minutes, has the lowest variance, 88.73 minutes, Problem 3 (HANNAH) has the highest mean and variance, and Problem 1 (DONALD) falls between these on both measures. On the other hand, we also observe that as the problem becomes more difficult, the relationship of subgroup variance is altered: in Problem 2, the verbal group has the lower variance, while the reverse is true in Problem 1, and there is no difference between the subgroups on Problem 3. However, before probing more deeply into this apparent relationship between problem difficulty and the effect of verbalization, it is necessary to examine the possibility that other factors may have been responsible for the observed variances.

It is often the case that uncontrolled factors affect the results of an experiment, and that these factors are not easily measured. In the present study, such factors appear to have had an influence on the results, but fortunately, some of these can be isolated for further examination. It is not intended to exaggerate the value of the analyses that follow, since the experiments were only designed to test the effects of verbalization, but it appears that the factors of

age and previous experience with cryptoarithmetic problems are independent influences on both length and variance of solution times. Such literature as exists on the influence of age and experience suggests that these factors have complex effects on performance. Hunter (1959) found age to be an important variable in the solution of anagram problems, with students superior to 15-year old school children, who were in turn superior to 12-year olds. However, he interprets these results in terms of the linguistic knowledge built up through an individual's experience with language, in particular knowledge of the statistical characteristics of relationships between letters, and this makes his findings essentially irrelevant for anticipating the effect of age in solving cryptoarithmetic problems. An additional finding indicated that there was no positive evidence of practice effects within the lists of the 36 anagrams used in the experiments. Similar conclusions can be drawn from a study confined to adult subjects solving anagrams had no effect on mean time to solution, but led to decreased variability among those with the most practice (Nance and Sinnot, 1963). It is suggested here that future experiments using cryptoarithmetic problems develop these themes more closely. Of particular interest would be whether it is age <u>per se</u>, or amount of mathematical experience that is important. Also,

various levels of cryptoarithmetic practice might be compared to Hunter's findings in relation to anagrams. It may be that some intellectual skills lend themselves to immediate improvement, whereas others improve only after prolonged practice (Means, 1967).

(b) The effects of age and previous experience

t-tests performed on the mean times for the adults and students showed significant differences between these groups on both problems. When the age groups were combined, and the effects of experience investigated, the inexperienced group had significantly higher mean time and variance on Problem 1 but not Problem 3. The tables containing these results are set out in Appendix B. Tests of significance were not performed on the data in Tables 4.2 - 4.5 because the number of cases being analyzed was too small in all instances.

Tables 4.2 and 4.3 present, separately for Problems 1 and 3, the means and variances for the adult and student groups according to whether or not there was previous experience with cryptoarithmetic problems. For this analysis, the subjects were divided into those who were in high school or had just graduated from high school, and the rest. Data is available for the two addition problems only, since the multiplication problem was not given to the student subjects (see Chapter 3 above, p.83).

Subjects were asked whether they had previously solved or tried to solve cryptoarithmetic problems. A subject was placed in the 'inexperienced' group for both problems if his familiarity derived solely from his experience with the first experimental problem. It is conceivable that any previous experience with these problems is the critical variable, in which case it would be unimportant whether this experience was very recent or whether it occurred months or even years earlier. However, the experiment was not designed to test this question, and further subdivision of subjects would result in cell numbers being too small for any analysis.

The results presented in Tables 4.2 and 4.3 indicate that age is the more important determinant in the proficient solving of cryptoarithmetic problems. Not only were the solution times obtained by the student groups consistently higher than those of the comparable adult groups, but, with one exception, the experienced students had higher mean times than the inexperienced adults.

TABLE 4.2

Mean Solution Times (minutes) and Variance Within Experimental Conditions for Adults and Students According to Previous Experience: Problem 1 (DONALD)

Age	Experience	Verbal			Nonverbal		
		N	Mean Time	Variance	N	Mean Time	Variance
Adult	Experienced	12	14.54	35.14	7	14.86	110.84
	Inexperienced	11	18.30	92.00	9	21.38	87.64
Student	Experienced	9	39.47	113.90	6	17.39	71.36
	Inexperienced	9	47.57	260.03	31	33.20	162.26

TABLE 4.3

Mean Solution Times (minutes) and Variance Within Experimental Conditions for Adults and Students According to Previous Experience: Problem 3 (HANNAH)

Age	Experience	Verbal			Nonverbal		
		N	Mean Time	Variance	N	Mean Time	Variance
Adult	Experienced	10	29.32	181.42	7	22.76	307.94
	Inexperienced	10	31.93	279.76	13	37.84	262.43
Student	Experienced	6	44.92	353.96	7	38.20	201.47
	Inexperienced	3	38.39	304.29	11	44.24	171.57

The exception occurred in the Problem I (DONALD)[1] nonverbal case, but in view of the small numbers of subjects involved, it is impossible to speculate further. There is a consistent experience effect within the age groups as measured by mean times, but the variance effects are less systemmatic. Again, this may be attributable to the small number of cases in each group.

---

[1]Note, however, considerable decrease in variance between verbal and nonverbal student groups. It is possible that this is due to the homogeneous structure of the group rather than to the effects of verbalization. This group (Experiment II) consisted of Fifth Grade students, with equal mathematical ability and training, who were tested in classroom groups. The majority of these students (84%) were not familiar with cryptoarithmetic problems, while other groups were much more evenly divided in this respect. While the nonverbal group was homogeneous on the variables of problem familiarity, mathematical training and ability, the verbal student group (Experiment III) had, by contrast, a minimum of mathematical training as their lowest common denominator. This latter group was comparable to the more hetoerogenous adult groups, both verbal and nonverbal (Experiment I). The difficulty in obtaining volunteer subjects made it impossible to correct these factors, which were, in any case, outside the main direction of the study.

TABLE 4.4

Mean Solution Times (Minutes) for Experienced and Inexperienced Groups According to Verbalization on Three Problems (Adults Only)

Problem	Experienced				Inexperienced			
	N	Verbal	N	Nonverbal	N	Verbal	N	Nonverbal
1 (DONALD)	12	14.54[2]	7	14.86	11	18.30	9	20.38
2 (SIR)	10	18.60	7	18.10	8	23.99	8	22.65
3 (HANNAH)	10	29.32	7	22.76	10	31.93	13	37.84

TABLE 4.5

Variance of Solution Times (Minutes) for Experienced and Inexperienced Groups According to Verbalization Three Problems (Adults Only)

Problem	Experienced				Inexperienced			
	N	Verbal	N	Nonverbal	N	Verbal	N	Nonverbal
1 (DONALD)	12	35.14	7	110.84	11	92.00	9	87.64
2 (SIR)	10	76.64	7	71.45	8	47.59	8	180.62
3 (HANNAH)	10	181.42	7	307.94	10	279.76	14	262.43

[2]Note that when the students are excluded from the analysis, Problem 1 replaces Problem 2 as the easiest in terms of mean solution time.

In view of the small numbers in each of the student subgroups,[3] it was not possible to perform any statistical analysis. Since age effects have been shown to be important, (Appendix B) and since the subgroup experience effects cannot be measured because of the small number of cases in each group, the students will be excluded from the next stage of analysis. Difference in the internal structure of the verbal and nonverbal student groups make it inadvisable to analyse the effects of experience within them.

By restricting the next analysis to adult groups only, it is possible to include the results of the multiplication problem for comparison. Although some of the very small cell sizes have been eliminated, the discussion that follows is still necessarily speculative.

The results for the adult groups are presented in Table 4.4 (mean times) and Table 4.5 (variances). The mean times for Problems 1 and 2 show the same tendency: the process of verbalization has no effect on experienced subjects. For the inexperienced, however, no clear trend emerges. In Problem 3, verbalization raises the mean time

---

[3] Although there were 31 subjects in the Inexperienced/Student/Nonverbal group (Table 4.2) the disparity with other subgroup sizes discouraged further analysis.

for the experienced, but lowers it for the inexperienced subject. This suggests that, for the able problem solver on a difficult problem, verbalization is a hindrance. On easier problems however, it has no effect. The less proficient, or inexperienced subject, is proportionally less affected by verbalization, and may even be helped by it. It is to be expected that mean times for the inexperienced group would be higher than for the experienced group, and this is borne out by the figures for Problems 1 and 2. On the other hand, for Problem 3, the absolute difference between the two verbal groups is much smaller, and between the two nonverbal groups much larger than for the other two problems, which suggests that the effect of verbalization should not be investigated independently of the level of problem difficulty.

The results of Table 4.5 indicate a similarity in performance on Problems 1 and 3: with verbalization, the variance is lower for the experienced than for the inexperienced group. Without verbalization however, the direction is reversed, and the experienced group has the higher variance. For Problem 2, the trend follows the opposite direction. An alternative formulation of the results would be that for Problems 1 and 3, verbalization lowers the variance for the experienced subjects, but makes no difference to the variance for the inexperienced groups.

On Problem 2, verbalization does not affect the variance of the experienced groups, but acts to lower the variance for the inexperienced.

No generalizations are possible, but it appears that experience and inexperienced subjects are differentially affected by having to verbalize, and that this effect is more pronounced in the most difficult problem.

(c) Relation between performance on Problems 1 and 3

Since the numbers of those subjects who had solved all three problems were too small for reliable analysis, correlation measures were performed on those subjects who had solved the two addition problems. Using Pearson's product moment correlation technique, the following results were obtained:

TABLE 4.6

Product-moment Correlations of Mean Solution Times Between Problems 1 (DONALD) and 3 (HANNAH)

Problem	Condition	N	r_{xy}
1-3 (DONALD-HANNAH)	Verbal	27	0.48**
1-3 (DONALD-HANNAH)	Nonverbal	17	0.42
1-3 (DONALD-HANNAH)	Verbal and Nonverbal	44	0.46**

** $p < .01$

There is a consistent degree of correlation between performance on the two problems, regardless of whether the subjects verbalized or not. This implies that an individual may have a personal rate of solution (λ) which is manifested on any problem attempted by him.

The mean times to solution on the two problems by the same individuals are significantly different, as shown in Table 4.7. The differences in variance are not significant however, which supports the hypothesis of individual solving characteristics being transferred from problem to problem.

TABLE 4.7

Mean Solution Times (Minutes) and Variance Under Verbal and Nonverbal Conditions Among Subjects Who Solved Problem 1 (DONALD) and 3 (HANNAH)

Condition	Problem	N	Mean Time	Variance	t (related means)	F
Verbal	1 (DONALD)	27	21.14	163.66		
	3 (HANNAH)		33.73	267.13	2.57*	1.63
Nonverbal	1 (DONALD)	17	17.07	86.34		
	3 (HANNAH)		28.36	190.88	4.27***	2.21

*p < .05
***p < .001

(d) Sex differences

Since the subjects were not sampled for the experiments, and since it was difficult to recruit volunteers, it was not possible to control for sex differences. Very few female subjects were used in the experiments, and it is impossible to

attribute any affect to them. The literature on sex differences in problem solving contains few studies in which complex mathematical problems have been used. Maccoby (1966), reviewing research on sex differences in intellectual functioning, cites studies with high school students which indicate that, on tests of arithmetic reasoning, boys show fairly consistent superiority. During college, men are superior to women in solving difficult methematical problems, especially when a 'restructuring' of the data is involved (Sweeney, 1953; Nakamura, 1955).

4.2 Estimates of K and λ: $\hat{k}$ and $\hat{\lambda}$

The second stage of analysis of solution times involved estimating the two parameters, K and λ, from the solution times data. Restle and Davis (1962) used the method of moments instead of maximum likelihood or other estimates, claiming it to be insensitive to experimental inadequacies. However, latency data frequently contain a few very large values, to which moments estimators would be sensitive. J.H. Davis (1964) examines the possibility of using other estimators. Moran (1957) considers the method of moments to be a highly inefficient procedure, and advocates a maximum likelihood estimator as the method which is asymptotically the most efficient. Although a maximum likelihood estimator is often biassed, it is to be preferred to other methods: the bias can

usually be adjusted by a constant which depends on sample size; it has asymptotic normality properties, and it has a minimum of variance. For a large sample, the asymptotic variance of a maximum likelihood estimator would be less than or equal to the variance of any other biassed estimator, while the variance of the method of moments is very difficult to calculate(Kendall and Stewart, 1958). A discussion of the principles of maximum likelihood estimation is presented by W.L.Hayes (1964, p.212-215). A more formal approach to estimation in general, and maximum likelihood methods in particular is adopted by Bush (1963, p.429-469).
Using the formula derived by Moran (1957), K and λ are estimated by choosing them to maximize the likelihood of the observations. The estimators, $\hat{k}$ and $\hat{\lambda}$, are the solutions of

$$-n \frac{\Gamma^{1}(k)}{\Gamma(k)} + \Sigma \ln x_i - n \ln \frac{k}{\bar{x}} = 0 \quad \text{(Eq. 4.1)}$$

$$\lambda^{2} \Sigma x_i - nk\lambda = 0 \quad \text{(Eq.4.2)}$$

From (Eq.4.1) we see that $\bar{x}=\hat{k}/\hat{\lambda}$ and we can solve (Eq.4.1) by successive approximation, choosing $\hat{k}$ first, and then $\hat{\lambda} = \hat{k}/\bar{x}$.

Table 4.8 presents the results of estimating K and λ using the method of maximum likelihood, with $\hat{k}$

and $\hat{\lambda}$ values given for verbal, nonverbal and pooled groups. Tests of significance were performed on differences between the verbal and nonverbal sample estimates, and, in later tables, between sample and theoretically determined values.[4]

Table 4.9 presents the values for $\hat{k}$ and $\hat{\lambda}$ when these are estimated by the method of moments. These results are presented for contrast with those obtained by maximum likelihood methods, and to provide comparison with the Restle and Davis data. Confidence limits for a difference of 1 between sample values of $\hat{k}$ were established, and the differences between verbal and nonverbal groups were evaluated in the light of these intervals.

The sampling distribution of differences of K and $\hat{\lambda}$ were calculated by Monte Carlo methods, and are discussed in greater detail in Chapter 5. Tables containing other values of differences of K and λ are presented in Appendix F.

---

[4] Normal curve methods were used to evaluate the significance of the results. The sampling distribution of any maximum likelihood estimator approaches the normal distribution as the sample size increases (Fisher, 1922). Caution must be exercised in the interpretation of estimates based on samples of size 30 or less, but with samples considerably larger than this, the normal distributon assumption is satisfactory.

TABLE 4.8

Values of $\hat{k}$ and $\hat{\lambda}$ For Three Problems as Estimated by the Method of Maximum Likelihood

Problem	Condition	N	$\hat{k}$	z	$\hat{\lambda}$	z
1 (DONALD)	Verbal	41	2.3438		0.0891	
	Nonverbal	53	2.5781	1.00	0.0946	.01
	Verbal and Nonverbal	94	2.5000		0.0931	
2 (SIR)	Verbal	16	3.4375		0.1671	
	Nonverbal	18	2.4219	2.83**	0.1218	.02
	Verbal and Nonverbal	34	2.7806		0.1428	
3 (HANNAH)	Verbal	29	2.8906		0.0841	
	Nonverbal	38	2.7344	0.48	0.0743	.01
	Verbal and Nonverbal	67	2.8125		0.0787	

** $p < .01$

The only significant difference between subgroup values for either $\hat{k}$ or $\hat{\lambda}$ appeared between the verbal and nonverbal groups on the SIR problem. In view of the small number of cases in these groups, this result should be interpreted with caution. The $\hat{k}$ and $\hat{\lambda}$ values for the other problems indicate that verbalization does not affect the stage structure of the problem solving solution process, nor does it influence the probability rate.

TABLE 4.9

Values of $\hat{k}$ and $\hat{\lambda}$ For Three Problems as Estimated by the Method of Moments

Problem	Condition	N	$\hat{k}$	$\hat{k}_{NV}-\hat{k}_{V}$	$\hat{\lambda}$	$\hat{\lambda}_{NV}-\hat{\lambda}_{V}$
1 (DONALD)	Verbal	41	2.5403		0.0965	
	Nonverbal	53	3.9216	1.3813*	0.1453	0.0488
	Verbal and Nonverbal	94	3.2215		0.1207	
2 (SIR)	Verbal	16	6.5309		0.3166	
	Nonverbal	18	3.2652	-3.2657**	0.1642	0.1524*
	Verbal and Nonverbal	34	4.6330		0.2285	
3 (HANNAH)	Verbal	29	4.3812	0.6455	0.1274	
	Nonverbal	38	5.0267		0.1364	0.0100
		67	4.8518		0.1368	

* $\hat{k}_{NV}=\hat{k}_{V}$ or $\hat{\lambda}_{NV}-\hat{\lambda}_{V}>1$

** $\hat{k}_{NV}-\hat{k}_{V}>3$

In all cases, the method of moments yields higher $\hat{k}$ estimates than does the maximum likelihood method. The difference of 1.3813 between the $\hat{k}$ values of the verbal and nonverbal groups on Problem 1 (DONALD) falls outside the 95% confidence limits of the mean of the distribution of $K_1-K_2=1$, which means that the $\hat{k}$ value for the nonverbal group is significantly larger, by more than 1. On Problem 2 (SIR), the verbal and nonverbal $\hat{k}$ values are significantly different by more than 3, whereas the verbal and nonverbal groups on Problem 3 are not significantly different. The differences in $\hat{\lambda}$ values are not significant for the two addition problems, but the verbal group solving Problem 2 had a significantly higher $\hat{\lambda}$ than the nonverbal group. Despite differences between the groups, the data was pooled to provide more reliable

estimates for each problem.

The range of $\hat{k}$ values is much larger when they are estimated by the method of moments, indicating that the variance of this estimator may be considerably larger than for the method of maximum likelihood estimation. This may also explain why the differences between subgroups were significant on two of the three problems when moments estimators were used.

The $\hat{\lambda}$ values in Table 4.9 are, at first sight, puzzling. Rate of solution is hypothesized to be lower with increasing module or problem complexity. However, in the method of moments estimates, the values are lower in Problem 1 (DONALD) than they are for Problem 3 (HANNAH), which was, using mean time to solution as an index, the more difficult problem. Problem 2 (SIR), using the same criterion, was the easiest of the three problems, and this is supported by the high $\hat{\lambda}$ value ($\hat{\lambda}$ pooled verbal and nonverbal = 0.2285). The explanation for the unsystemmatic behaviour of the $\hat{\lambda}$ values lies in the fact that the method of moments estimates $\hat{\lambda}$ as the ratio between mean and standard deviation. On Problem 1, the relatively high variance in relation to the mean (see Table 4.1) resulting in a low value of $\hat{\lambda}$. On Problem 3, the variance was lower relative to the mean, which yielded the surprisingly higher $\hat{\lambda}$ estimate.

On the other hand, the maximum likelihood estimates of (Table 4.8) support the hypothesis that the value of $\hat{\lambda}$ is

inversely related to problem difficulty. Problem 2 is associated with the highest value ($\hat{\lambda}$ = 0.1428), followed by Problem 1 ($\hat{\lambda}$ = 0.0931) and Problem 3 with the lowest value ($\hat{\lambda}$ = 0.0787) which is consistent with the ordering of difficulty of the problems by mean time to solution (Table 4.1).

Both methods of analysis are consistent in that a higher $\hat{k}$ is estimated for both subgroups and the pooled group on Problem 3 than on Problem 1, which suggests that the former problem is indeed the more complex. The relative position of the $\hat{k}$ estimate for Problem 2 is inconsistent according to the estimator used, but since the values are based on a small number of cases, it is idle to speculate on reasons for this phenomenon.

4.3 Fitting the Gamma Distribution to the Empirical Distributions

Cumulative relative frequency distribution of obtained solution times were fitted by the appropriate gamma distributions for the verbal, nonverbal and pooled verbal-nonverbal groups in each problem. The graphs showing these fits are presented in Figures 4.1 to 4.3. The points of the theoretical curves were generated by computer, and provided acceptable fits to the data for all three problems, when goodness-of-fit was tested using the Kolmogorov-Smirnov one-sample test (Siegel, 1956). The largest discrepancy

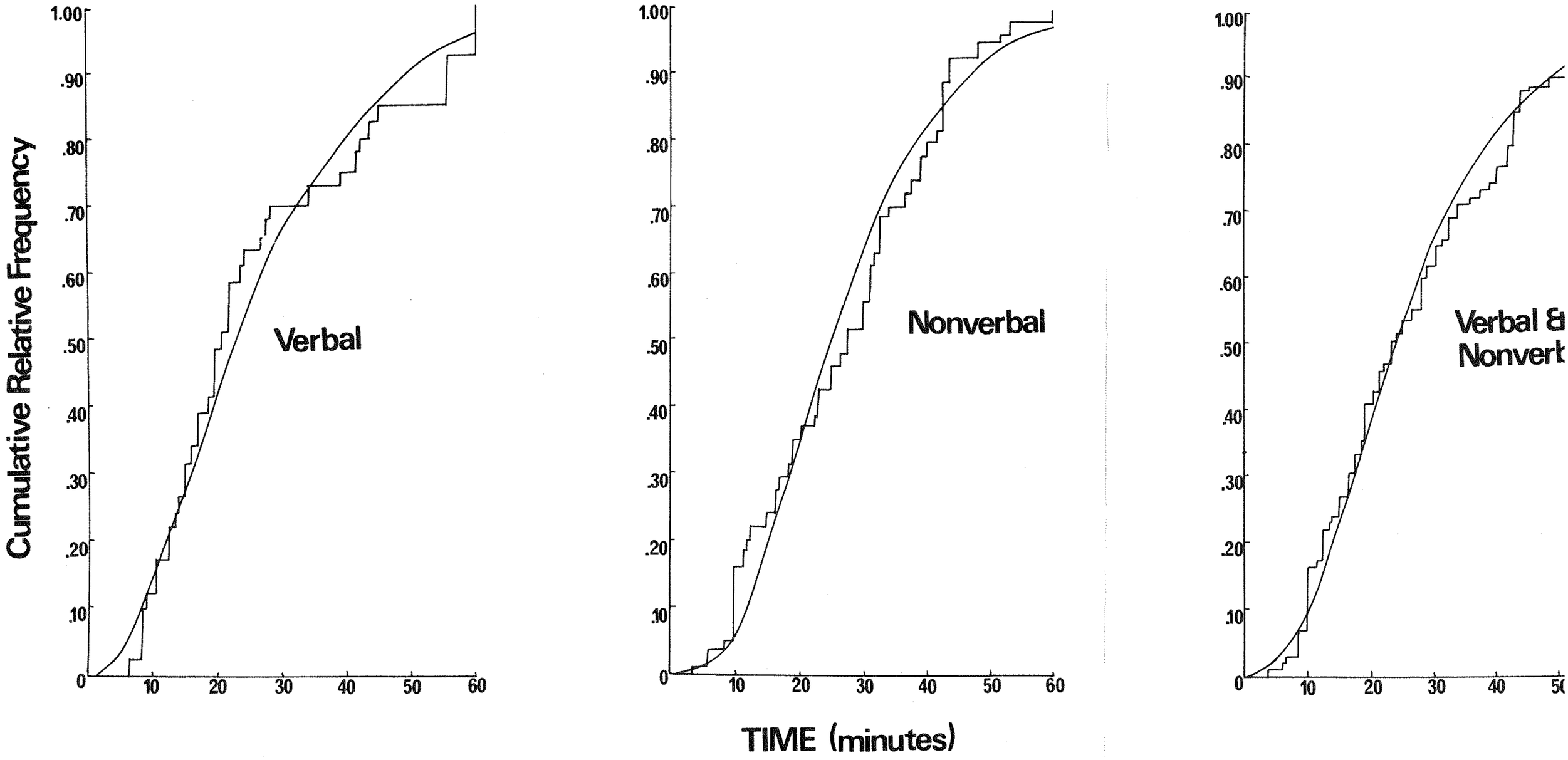


FIGURE 4.1: Problem 1 (DONALD) Cumulative relative frequency distributions of solution times (step function) with fitted cumulative gamma distribution (smooth curve).

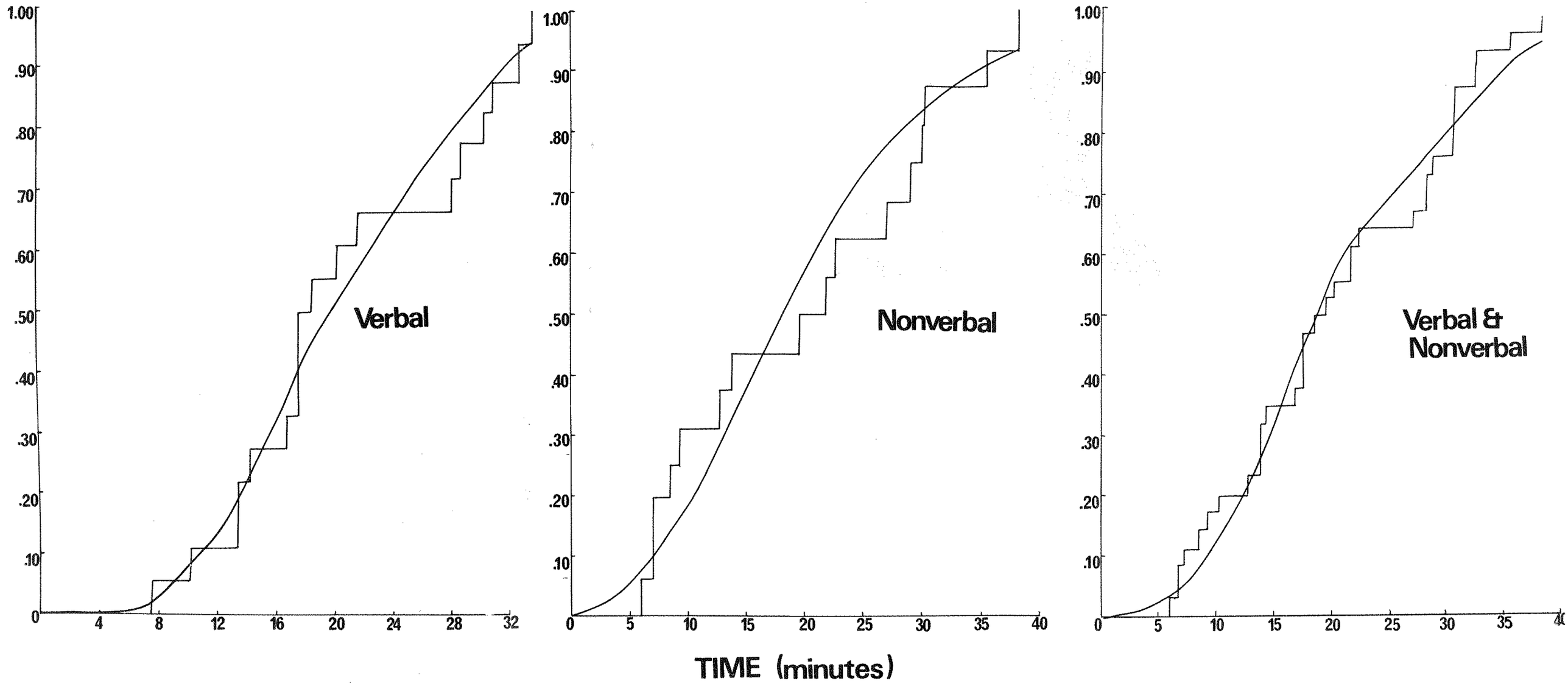


FIGURE 4.2: Problem 2 (SIR) Cumulative relative frequency distribution of solution times (step function) with fitted cumulative gamma distribution (smooth curve)

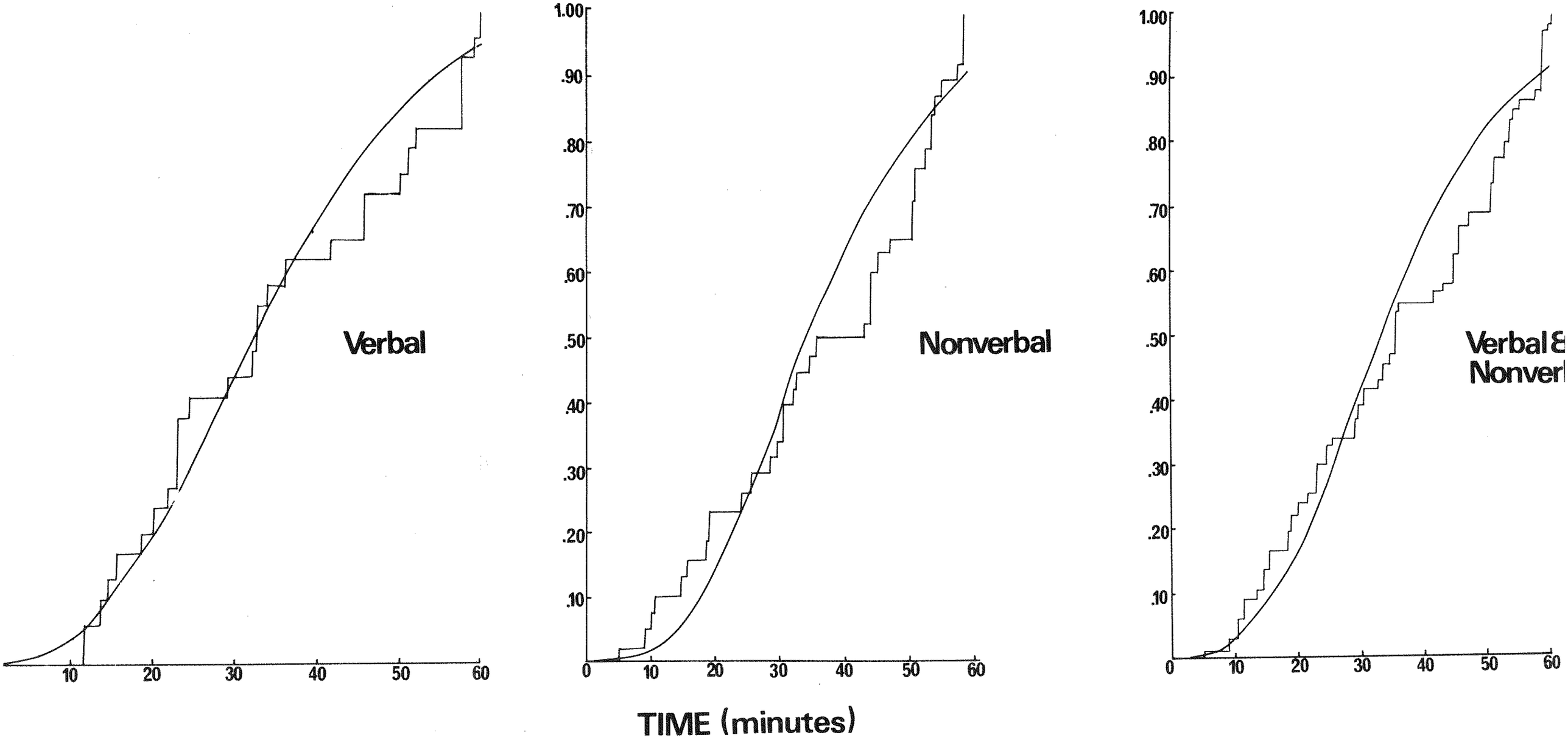


FIGURE 4.3: Problem 3 (HANNAH) Cumulative relative frequency distribution of solution times (step function) with fitted cumulative gamma distribution (smooth curve).

between theoretical and empirical distribution is usually found in the top half of the distribution. The points at which the difference between the two distributions is at a maximum, along with the size of that difference are presented in Appendix C. Notice that the cumulative relative frequency of the theoretical distribution approaches, but does not reach 1.00. This is partly influenced by the fact that the empirical distributions are truncated,since no subject was allowed to spend more than sixty minutes on a problem. The question of truncation is discussed at the end of the chapter (p.139).

4.4 Analyzing the Protocols for Stages

(a) Rationale

If, as seems reasonable, the definition of a module must be anchored to the physical structure of a problem, then this definition must include either an individual letter or a group of letters. The protocols produced by the subjects in the verbal groups were analyzed to provide independent verification for the number of stages estimated by the solution times. Analysis consisted of examining the protocols in order to specify the sequence in which each subject correctly assigned digits to the letters of the problem. When this was completed, the time to each correct assignment was measured, using the method described in Chapter 3. There was occasional ambiguity concerning the order in which letters were decoded.

This could occur if the subject did not make explicit his substitution of a digit for a letter, or if a digit that had been correctly assigned was later either 'forgotten' or changed. Rules were established to maximize a consistent interpretation of such situations. Definite substitution was considered to have occurred the first time the subject categorically asserted that it had. For example, the statement "R is definitely 7" constituted such an assertion, whereas "R may be 7, could be 9, I'm not sure", did not. If no more definite statement occurred later in the protocol, then statements of the latter type were taken as evidence of substitution. In the absence of this kind of statement, the protocol was examined for evidence in some other form, as for example, the correct substitution occurring during a series of computations. It was rarely necessary to resort to this last level of evidence. Subjects clearly differed in their willingness to make categorical statements, no matter how certain they were about the correctness of a hypothesis or an assignment.

Decision rules were more difficult to apply in the case where a subject changed his mind or forgot a substitution he had previously made correctly. If he made a categorical, correct statement concerning a particular letter, but disregarded the information in his subsequent processing, it was not considered as an assignment. If he used the information, but then changed his mind, reversing his line of

reasoning, this was not considered as an assignment until he correctly resumed his position, and used the implications of this reasoning. These rules have been set out in some detail, since the rationale according to which they were developed is central to an investigation of the stages-module model.

In the first step of analysis, the time to solve each letter was calculated, along with its $\hat{k}$ and $\hat{\lambda}$ values. These individual letter distributions did not have the properties necessary to qualify for a stage: that is, the $\hat{k}$ values were not approximately equal to 1. In the second step, letters were combined into groups, according to the sequence in which they were converted into digits. It was possible to discern an invariable sequence in which groups of letters were decoded for almost all subjects, although there was variability of letter sequence within the groups. This will be discussed in much greater detail in Chapter 6. The protocols taken from subjects in the verbal groups, and supported by the post hoc reconstruction of sequences by the nonverbal groups, suggested that, for the problems to be successfully solved, it was necessary for certain letters to be decoded before further progress could be made.

A stage has been defined as an exponentially distributed random variable describing the time to solve a module. In an exponential distribution, the mean is equal to the standard

deviation, and $m^2/s^2=1$. If the distribution of times to complete a module (i.e. to solve a given group of letter-digit substitutions) is in fact exponential, then estimates of K should be approximately equal to 1 for that stage, and the empirical distribution of solution times should be fitted by the appropriate exponential distribution. The next section will describe the results of estimating the stage values from the protocols.

At this point of the study, we can relate the module structure of the problem, which has been arrived at independently by examination of the protocols, to two separate estimates of stages:

(1) the value of $\hat{k}$ estimated from solution times to solve a given module, and

(2) the fitting of an exponential curve, using the appropriate parameter (λ) value, to the cumulative empirical distribution of solution times.

In this chapter, the stages and modules are simply referred to by number. Their qualitative structure will be discussed in Chapter 6. The distinctions and relationships between stages and modules were discussed in Chapter 2.3.

(b) Stage estimates

Analysis of protocols was restricted to the two addition problems. The analysis indicated that Problem 3 (HANNAH) had three stages, and that Problem 1 (DONALD) had

four. The $\hat{k}$ values for each stage were calculated by maximum likelihood estimators. In no case was the value for a stage significantly different from 1. Table 4.12 presents the $\hat{k}$ and $\hat{\lambda}$ values for the two problems, along with the mean solution time and variance of each stage.

TABLE 4.10

Mean Solution Times (Minutes) and Variance For Stages of Problem 1 (DONALD) and Problem 3 (HANNAH) and $\hat{k}$ and $\hat{\lambda}$ Values for Each Stage Estimated by Method of Maximum Likelihood

Problem	Stage	N	Mean Time	Variance	$\hat{k}$	$\hat{\lambda}$
1 (DONALD)	1	40[5]	1.06	0.93	1.2500	1.1821
	2		8.75	7.55	1.4425	0.1785
	3		12.24	12.24	1.2188	0.0955
	4		3.06	2.83	1.3125	0.4285
3 (HANNAH)	1	29	8.05	7.89	1.1250	0.1398
	2		22.62	15.86	1.7188	0.0760
	3		3.09	3.10	1.3750	0.4450

The $\hat{k}$ value for HANNAH Stage 2 is closer to 2 than to 1[6] but all the other values approximate 1. The reasons for this will be discussed in Chapters 6 and 7, in the context of a closer analysis of problem structure and solution process.

[5] The tape recording of one subject could not be used.

[6] But not significantly different from either: Presumably with a larger sample size, the value $\hat{k} = 1.7188$ would differ significantly from 1. Again, a note of caution is introduced about maximum likelihood estimation and adequate sample sizes.

For comparison, the $\hat{k}$ and $\hat{\lambda}$ values estimated by the method of moments are presented in Table 4.11. The $\hat{k}$ values approximate 1 more closely than do those of Table 4.10. In contrast to the estimates for the total problem (Tables 4.8 and 4.9), the individual stage values are lower when calculated by this method than by the maximum likelihood method. Notice that this is not true of Stage 2 in Problem 3 where the moments estimate is higher than the maximum likelihood estimate. It appears that the relative direction of bias of the two methods cross over somewhere between $\hat{k} = 1$ and $\hat{k} = 2$.

TABLE 4.11

$\hat{k}$ and $\hat{\lambda}$ Values for Problem 1 (DONALD) and Problem 3 (HANNAH) Estimated by the Method of Moments.

Problem	Stage	N	$\hat{k}$	$\hat{\lambda}$
1 (DONALD)	1	40	1.1963	0.1292
	2		1.3450	0.1142
	3		1.0854	0.0854
	4		1.1739	0.3832
3 (HANNAH)	1	29	1.0401	0.1292
	2		1.9921	0.0880
	3		0.9927	0.3212

The 95% confidence interval for K=1, when N=40 is 0.54-1.78, and when N=29, 0.54-2.12. The 95% confidence interval for K=2, when N=29 is 1.10-3.30.

The $\hat{k}$ values, with the exception again of Stage 2 in Problem 3 are very close to 1, and certainly all fall within the 95% confidence interval for K=1. Most of them also fall within the limits for K=2, and this considerable degree of overlap between sampling distributions presents a problem in the interpretation of the statistics. The sampling distribution of $\hat{k}$ does become tighter as $\hat{k}$ becomes larger, but in the discussion of individual stages, we are dealing with the sampling distribution when $\hat{k}=1$. In the absence of more rigorous evaluation procedures, it is reasonable to consider the estimates in terms of their nearest integer value. This means that all the stages we have observed pass the test stipulated by the theory, except for the second stage in Problem 3, which is better considered as being composed of two stages.

There is considerable variation in the $\hat{\lambda}$ values estimated for each stage. The maximum likelihood exceed the moments estimates with the exception of Stage 2 in Problem 3. The ordering of the values is the same for either method of estimation. The middle stages in both problems appear to have been the most difficult, that is, they have the smallest $\hat{\lambda}$ values. The exceptionally high rate ($\hat{\lambda}$ =1.1281) for Stage 1 (DONALD) is explained by the fact that one of the letter assignments of that module was given to the subjects prior to starting on the problem, and the

second letter followed immediately from that. Notice (Table 4.10)that the mean times and $\hat{\lambda}$ for the last module in either problem are remarkably similar. This will be referred to again in Chapter 7. It is clear that the modules are not equal in difficulty whether the index used is mean time or $\hat{\lambda}$.

(c) Independence of stages

The theory of waiting times implies that the modules of a problem are solved independently, since the stage times are described by independent random variables. To determine whether this assumption was violated in the experiment, Pearson product moment correlation tests were performed on the data. The results are given in Tables 4.12 and 4.13.

TABLE 4.12.

Correlation Matrix Between Stages of Problem 1 (DONALD)

	1	2	3	4
1		.44**	.31*	.17
2			.14	.31*
3				.24

* $p < .05$
** $p < .01$

TABLE 4.13

Correlation Matrix Between Stages of Problem 3 (HANNAH)

	2	3
1	-.09	-.03
2		.14

It is clear that there was some degree of dependence between the stage times taken to solve the modules of Problem 1, although there is no evidence of such a relationship for Problem 3. This implies that, for Problem 1, performance on Module 2 is related to performance on Modules 3 and 4, whereas the stages model would predict zero correlation between all modules. The results suggest that, on Problem 1, subjects tend to differ from each other in the same way over modules. In other words, the value on Stage 1, whether it be the value of the mean time or λ, determines to some extent the range of values possible on Stage 2.

Waiting times theory also states that the stages to solve each module are identically distributed, that is, with the same mean and λ. It is redundant to perform tests of significance between the means and variances of individual stages since they are clearly not identical, but display a wide range of values. This is equally true for either problem

(see Table 4.10). Clearly, the modules therefore differ to a significant degree in level of difficulty as measured by mean time taken to solve them, as well as by their associated λ values. The model's assumption of independence and identical parameter values are violated in these experiments. The violation of these assumptions may have been responsible for the inability of the model to reconcile the individual stages estimates and their sum (see Section 4.4e below). However the assumptions appear to be dispensable when fitting the gamma distribution to these data.

(d) Fitting the exponential distribution to the empirical distributions

The final test of the model lies in discovering whether the waiting time (stage) to solve each module is exponentially distributed. The cumulative distribution of solution times for each of the four stages of Problem 1 (DONALD) and the three stages of Problem 3 (HANNAH) were graphed, together with their appropriate exponential distributions. The results of these procedures are presented in Figures 4.4 and 4.5.[7] Goodness-of-fit tests were performed using

[7] Notice in Figure 4.5 that the shape of the distribution for Stage 2 is closer to that of the gamma distributions in Figures 4.1 to 4.3 than it is to the exponential distributions for the other stages in Figures 4.4 and 4.5. This is compatible with the finding that the value of $\hat{k}$ for Stage 2 in Problem 3 is closer to 2 than to 1, and with the fact that the gamma distribution reduces to the exponential distribution when K=1.

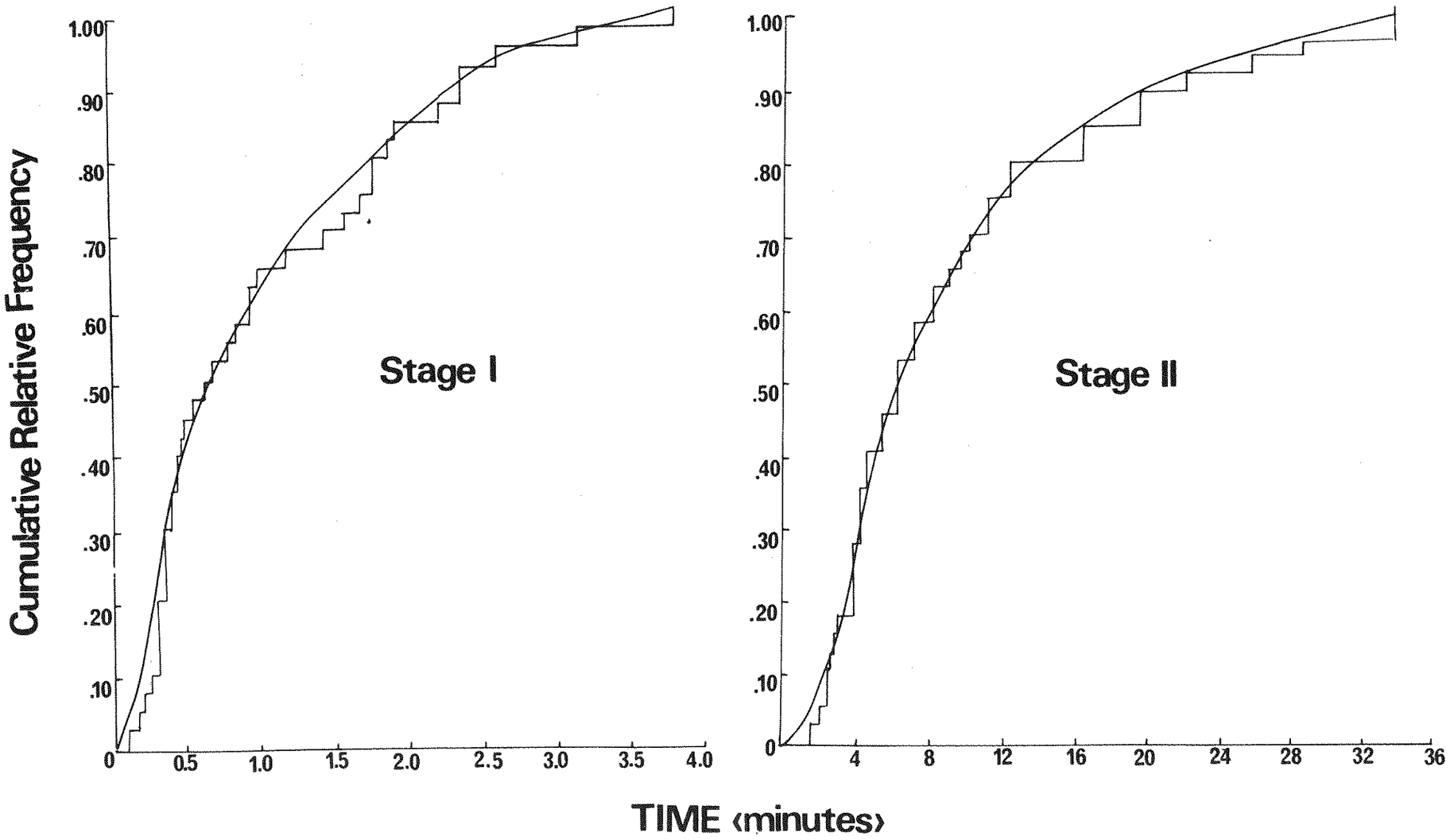


FIGURE 4.4: Problem 1 (DONALD) Cumulative relative frequency distributions of solution times for Stages 1 to 4 (step function) with fitted cumulative exponential distribution (smooth curve).

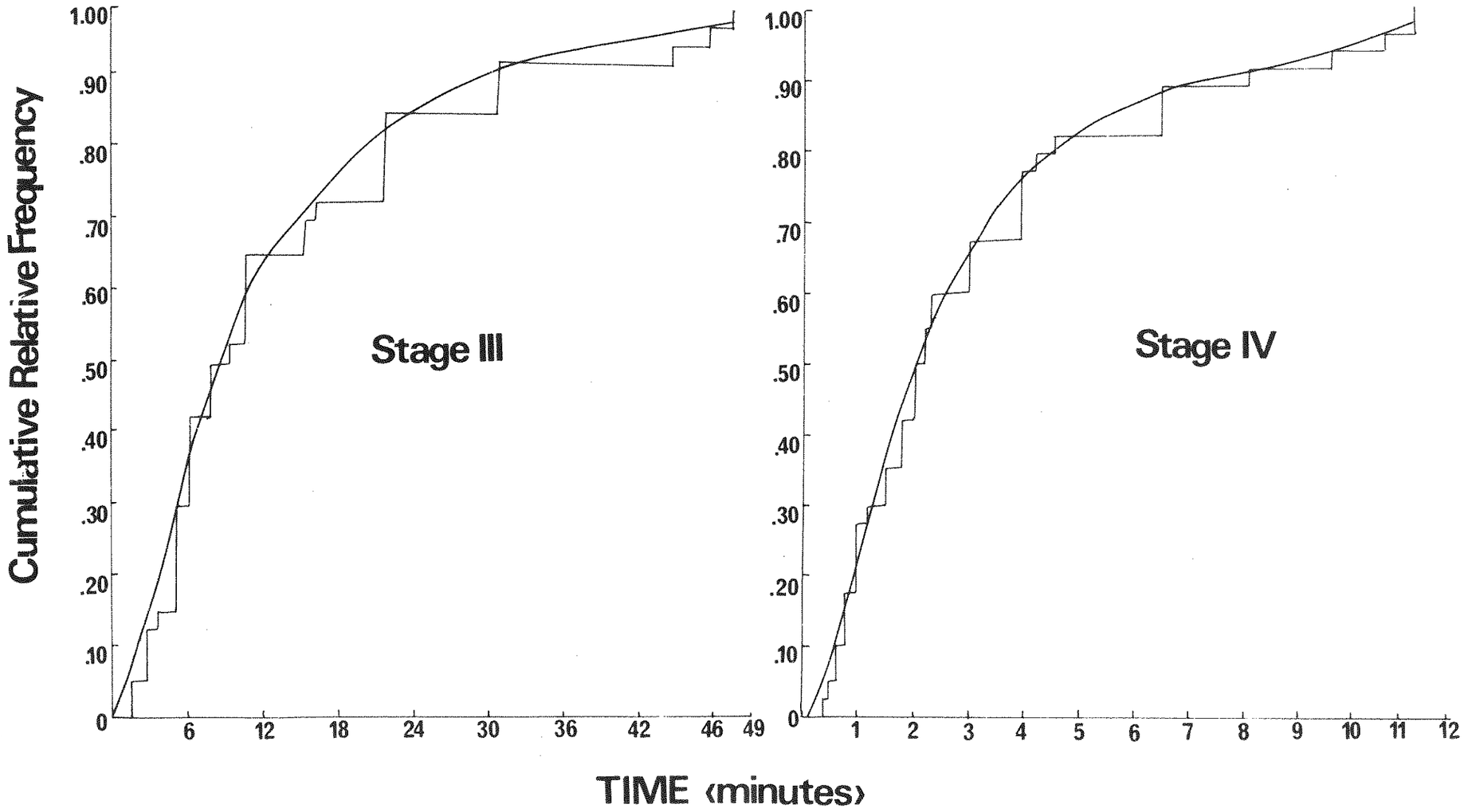


FIGURE 4.4 (Cont'd)

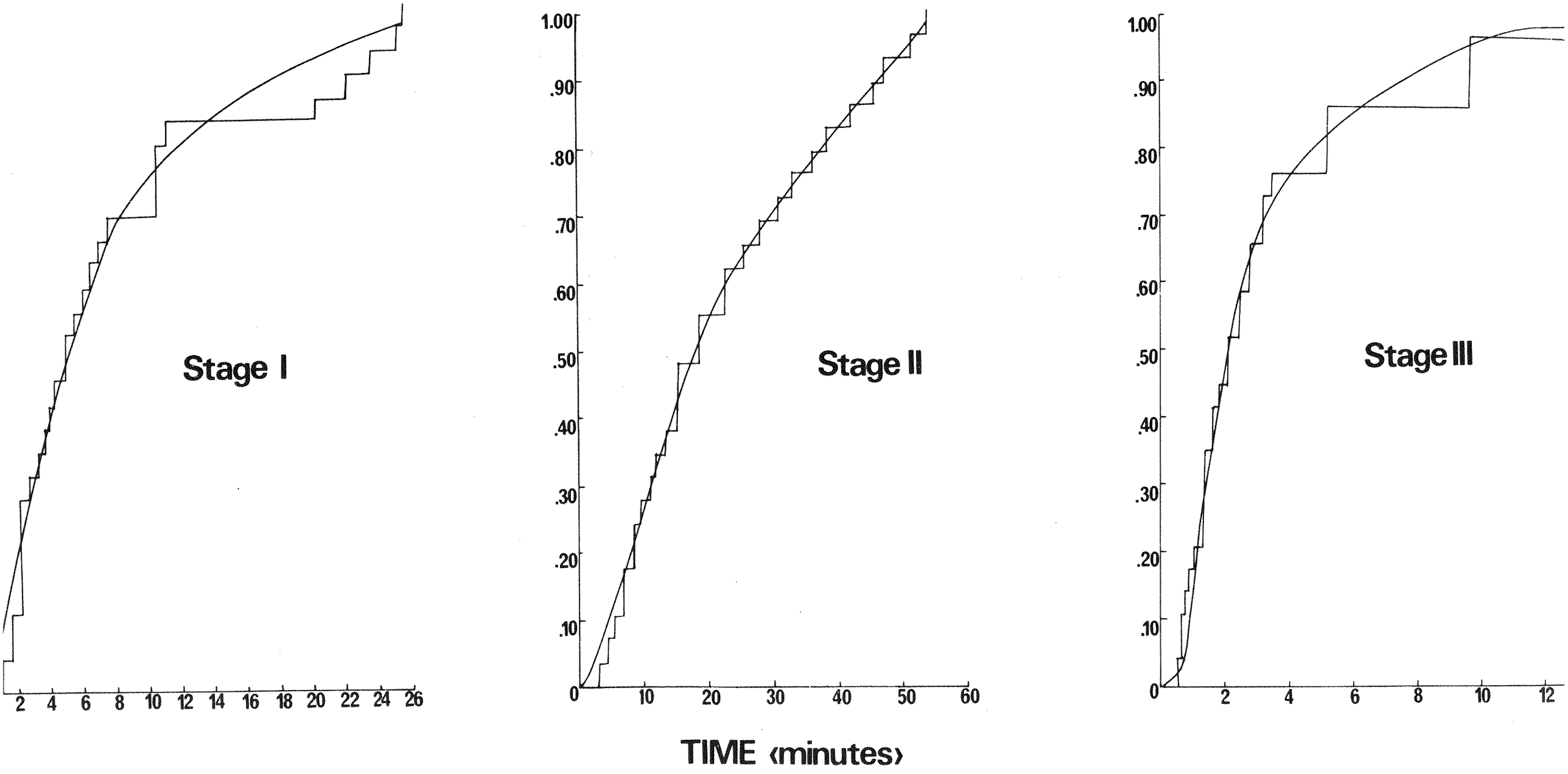


FIGURE 4.5: Problem 3 (HANNAH) Cumulative relative frequency distributions of solution times for Stages 1 to 3 (step function) with fitted cumulative exponential distribution (smooth curve)

the Kolmogorov-Smirnov one-sample test. The maximum differences between theoretical and empirical distributions supported the null hypothesis in all cases; the points and size of maximum difference are presented in Appendix C.

(e) Summing the stages

It only remains to assess the relationship between the individual stages and the number of stages estimated for the total problem. Theoretically, the number of individual stages should sum to the number estimated from the total solution times data for the whole problem. More specifically, how does the $\hat{k}$ estimate for Problem 1 compare with K=4 and for Problem 3 with K=3, the values derived from analyzing the protocols? Table 4.14 presents the pooled $\hat{k}$ values for each problem, using both methods of estimation.

TABLE 4.14.

$\hat{k}$ Values Estimated for Problem 1 (DONALD) and Problem 3 (HANNAH) Compared with the Number of Stages and Modules Isolated From the Protocols

Problem	N	$\hat{k}$ (estimated by method of maximum likelihood)	$\hat{k}$ (estimated by method of moments)	K (identified from protocols)
1 (DONALD)	94	2.5000	3.2215	4
3 (HANNAH)	67	2.8125	4.8518	3

The value of 2.5000 for Problem 1 is significantly different ($p = < .001$) from both 2 and 3 ($z=3.26$ and 3.46 respectively). Even if the value was considered in terms of its nearest integer equivalent, it would still be a poor estimate of the protocol value of 4. The moments estimate of 3.2215 falls into the confidence region for K=3 and K=4. It is unsatisfactory to place significance on this result when the confidence intervals are as wide as these, even with relatively large number of subjects (N=94). Selected 95% confidence intervals are presented in Table 4.15. A more complete range of values is found in Appendix D.

TABLE 4.15

95% Confidence Intervals For Sampling Distribution of $\hat{k}$ with Mean K=2-5

K Mean	N	95% Confidence Interval
3	94	2.1430-4.2562
4	94	2.8500-5.5889
3	67	2.0831-4.5000
4	67	2.6754-5.9150
5	67	5.3330-7.4400

The maximum likelihood estimate for Problem 3 is not significantly different from the protocol-based value of K. However, the pooled method of moments estimate is open to a large range of confirmation, and virtually any conclusion may be supported. Although the maximum likelihood

estimate appears to provide support for the stages-module model, closer analysis suggests that this estimate is in disagreement with the finding that the $\hat{k}$ value for the second stage in Problem 3 is closer to 2 than to 1, which means that the sum of the stages is closer to 4 than to 3.

Clearly, the stages-modules model does not find support from a more thorough inspection of the solution process. The $\hat{k}$ values, as used by Restle and Davis (1962) are purported to be estimates of processes hidden to the investigator in the absence of independent verification. When such verification is made possible through the method of protocol analysis, however, the values obtained by applying the theoretical model do not describe the situation. This is probably due in large part to the fact that the stages model breaks down when certain assumptions are violated. Since the separate exponentially described stages were not independent, and did not have identical parameters or means, the sum of these distributions is not gamma but some other unknown distribution. The $\hat{k}$ estimated for the total number of stages in the problem is, in the two problems under discussion, less than the sum of the component processes discovered through independent analysis. The fact that the solution times for both individual stages and the total problem are satisfactorily described by the appropriate exponential and gamma distribution respectively, indicates that the concept of

goodness-of-fit does not provide an adequate test of the model. It must be concluded that the descriptive flexibility of these distributions limits the sensitivity of the model since serious discrepancies and violations are obscured in the satisfactory fitting of the model to the data. The stages model makes several assumptions, and the more assumptions a model makes, the easier it may be to apply the resulting structure, but the less likely is it to correspond to the element or the situation modelled: Given a situation that is described by n+1 points, for example, it is always possible to fit exactly to those points a polynomial of the n^{th} degree, although this is more complicated than fitting a gamma distribution where only two parameters have to be calculated. Therefore, restraint must be exerted in overestimating the importance of a good fit of a theoretical to an empirical distribution.

4.5 Conclusion

It is apparent that all known analytic paths for assessing the data have been reviewed. Maximum likelihood estimators have been shown to be superior to the moments estimators used by previous investigators. However large samples are desirable if the results of significance tests are to be satisfactorily evaluated, and the calculation of confidence intervals through Monte Carlo procedures would be extremely expensive. The confidence intervals established

for the method of moments estimators are too broad for rigorous testing. It is not unreasonable in this situation simply to consider the obtained estimates in terms of their nearest integral value, but we have seen that this does not obviate the discrepancy between the products derived through statistical and protocol-based analytic procedures. It becomes conceivable that the stages model, unless it can be fundamentally retooled, may be useful only as a quantitative implement for guiding the analysis of protocols.

Note:

Since subjects were stopped if they had not solved the problem within sixty minutes, it is conceivable that the distribution of solvers may be truncated. However, since sixty minutes represents a considerable length of time to solve the particular problems used (the mean time to solution was at most thirty-five minutes for the most difficult problem), and since those subjects who had not solved by then did not appear to be close to solution, there is no reason to believe that extending the time before curtailing the subjects would have substantially changed the distributions. Further, the estimation of the parameters of a truncated gamma distribution, even by the method of moments, is tedious (Cohen, 1950). Chapman (1956) gives a new method of estimating the parameters by least square methods, but again, the procedure is lengthy, and not justified by the demands of the situation.

CHAPTER 5

MONTE CARLO SIMULATION STUDIES

The discussion at the end of Chapter 2 pointed to the need for simulation studies so that the behaviour of the stages-module model could be further explored and tested. In this chapter, a brief introduction to the techniques of Monte Carlo simulation (Section 5.1) is followed by an analysis of the data generated by these techniques. The simulation studies fall into two main categories, each designed to answer a distinct set of questions: the first category deals with sampling experiments that are used to generate sampling distributions of the parameters of the gamma distribution (K and λ). This is presented in Section 5.2. The second category of studies is designed to answer specific questions about the behaviour of the model when particular assumptions are violated. This is discussed in Section 5.3. The main body of the results of the simulation studies are presented in Appendix D-H. Only these results which are directly relevant to the development of the argument that follows will be found in the chapter itself.

5.1 The Monte Carlo Method

Computer simulation of the qualitative behaviour of single subject represents no different a process than that of

using the computer as a tool of numerical analysis in order to simulate the behaviour of a mathematical model. The model is first studied empirically, after which its behaviour is explored under a series of parameter variations. In Chapter 1, we reviewed some computer models of cognitive processes. In this chapter, we shall deal with computer simulation of mathematical models.

The stages-module model treats behaviour as a probabilistic, rather than as a deterministic process. More precisely, the underlying concept is that of a stochastic process, that is, a temporal sequence of events that can be interpreted in terms of probability theory. A stochastic model makes statements about the likelihood of certain events, and has built into it assumptions about the variability of behaviour. Monte Carlo simulation is usually used synonymously with 'simulation of stochastic processes', although in the past it was applied only to stochastic simulation for solving deterministic problems, which had proved intractable using standard numerical methods. Stochastic simulation methods are appropriate in situations where actual sampling would be either impossible or expensive. In other situations,information about expected values and moments may not provide an adequate description of a process, and Monte Carlo methods may supply the only

means for extracting the necessary information. Information can also be obtained about the variance of estimators by using a large number of Monte Carlo runs, and then applying the estimator to the data of each run.

A stochastic simulation differs from a classical sampling experiment since it requires the construction of a probabilistic model, whereas in a sampling experiment, operations are performed directly on the raw data.

Essentially, the Monte Carlo method is a technique for introducing data of a random or probabilistic nature into a model. The technique entails generating pseudo-data by precisely conforming to the rules given by the probability model (in this case, the gamma distribution), and consulting a random number generator whenever a certain, probabilistically determined, event occurs. There are specific techniques for generating the values of a random variable from probability distributions.[1] These, and other topics of computer simulation, are comprehensively treated by Naylor, Balintty, Burdick and Chu (1966).

---

[1] Basically, the problem of sampling from any distribution is that of transforming a random number representing the uniform random variable in the range (0,1) by means of, for example, the inverse cumulative function of the distribution of interest. Although the cumulative distribution function for the gamma distribution does not exist in explicit form, the values of the so called incomplete gamma function have been extensively tabulated by Pearson (1922).

In the present study, a standard computer library program was used to generate random numbers.[2] Strictly speaking, the processes used are not random at all, but strictly determinate, since digits are generated according to an arithmetical rule. However, these pseudorandom numbers may be considered random since they fulfill the requirements for sample values of a random variable: inter alia, they are statistically independent, they come from a uniform distribution where all numbers are equally likely, and they are nonrepeating for any desired length of series.

The pseudo-data from a Monte Carlo series of experiments indicates how 'real' data would behave if the parameters of the model (K and λ , in our case) had the particular values assumed in the computations. Among the earliest investigators to use Monte Carlo methods to test psychological models were Bush and Mosteller (1955), who coined the term 'stat-rat' to refer to the 'organisms' being run in the Monte Carlo experiments.

The essence of simulation experiments using the Monte Carlo method is that the extensive calculations produced are the results of relatively short calculations, each repeated a large number of times. Despite the inherent advantages of the method, therefore, it is also unfortunately

[2] For methods of providing random numbers for use on digital computers, see Teacher (1954).

true that, even with high speed computers, such simulation may be extremely costly, since millions of random digits and hours of computer time may be necessary to adequately test the features of a model.

5.2 Sampling Distribution Experiments

The data presented in this section were generated to investigate the form of the sampling distributions of K and λ, the two parameters of interest, since these distributions are unknown. More particularly, the object of these experiments was to determine the degree and form of the variability that the estimates of K and λ would have in repeated sampling. Knowing this, it would then be possible to make statements about the likelihood that a particular sample $\hat{k}$ or $\hat{\lambda}$ could have come from a distribution with a given mean K or λ .Alternatively, statements could be made concerning the probability that two samples could have come from a distribution with the same mean.

Since the method of moments was used to estimate the parameter values in the Monte Carlo experiments, the sampling distributions thus generated are valid only for interpreting the empirical values estimated by the method of moments, and may not be used to evaluate the values estimated by the method of maximum likelihood. Although the method of moments has been shown to be an unsatisfactory

estimation procedure (Chapter 4), the data produced in these sampling experiments, being the outcome of a large number of trials, provide reliable information about the behaviour of the parameters of the gamma distribution beyond that required for significance testing, and which is of independent interest and significance.

The sampling experiments fall into three main categories:

(1) The sampling distributions of $\hat{k}$ and $\hat{\lambda}$ were determined for a range of K from 1 to 9, with λ at .10 and .30 for each value of K. The sampling distributions were computed for various but not all combinations of K and N (sample size) where N ranged from 15 to 95. With the aid of frequency distributions, the tails (2½%) of the sampling distributions were graphed, so that tests of significance could be made on appropriate sample estimates. Confidence intervals for the mean of the sampling distribution were computed. Means and standard errors are available for each combination of K and N. These are presented in Appendix D. The effects on mean and standard error of varying the size of the sample for different values of K are shown in Appendix E.

(2) The sampling distributions were generated for differences of $\hat{k}$ and $\hat{\lambda}$ for some combinations of differences ranging from 1 to 7, and sample size varying from 15 to 92. Confidence intervals for the means of these distributions are presented in Appendix F, along with their means and standard errors.

(3) A further series of sampling experiments was performed to investigate the effects of variations in sample size from 18 to 100 on the sampling distribution of $\hat{k}$, when K=6, in order to observe the asymptotic normality properties of distribution, and thereby determine what would constitute an adequate sample size for the empirical investigation. The data from this series of Monte Carlo experiments is presented in Appendix G.

In most cases, each sampling distribution was the outcome of one thousand Monte Carlo trials, although in a few cases, only five hundred trials were performed.

5.3 Experiments to Investigate Effects of Violating Assumptions of The Stages-Module Model

In a previous chapter (2.3) the assumption of a constant conditional probability mechanism was discussed, along with its implications for individual differences, and for rate of solution on each module of a problem. In this context, it is possible to ask more specific questions.

(1) Given that individual differences exist, how much are subjects likely to differ in their rate of solution on a given module or problem? Is it reasonable, for example, for one subject to be considered two or three or four times 'better' than another subject?

(2) In an experiment, or a series of experiments, where λ is varied over a certain range, will the appropriate gamma distribution still satisfactorily describe the obtained (simulated) distribution of waiting times?

(3) What range of difficulty is it reasonable to hypothesize for the modules of a problem? Is it reasonable, for example, for one module to be two or three or four times more difficult than another module?

(4) What is the effect on the model, that is, how closely does it estimate the population parameter values, when the probability rates for the stages differ, but when each subject has the same probability rate on a given stage? Alternatively, how is the behaviour of the model affected when each subject has a constant probability rate on every module, but when the subjects differ in their probability rates?

(5) How is the behaviour of the model affected when λ is varied randomly (within a given range) and simultaneously for both subjects and stages?

(6) How does the size of K influence these situations?

The second category of Monte Carlo experiments were intended to answer the above questions. Specifically, since the stages-module model assumes a constant scale factor (λ) for subjects and stages, how is the model affected, i.e. how well does it estimate the parameters K and λ , when λ is varied

(1) for subjects, but kept constant for stages (VaSs)?

(2) for stages, but kept constant for subjects (VaSt)?

(3) for both subjects and stages (VaSsSt)?

How does the system respond to such violations when compared with its behaviour under the strict assumptions of the model, that is

(4) when λ is constant for both subjects and stages (CoSsSt)?

In Table 5.1, we show the four different conditions under which simulation of the behaviour of the model proceeded.

TABLE 5.1

Four Paradigms For Testing the Effects of Varying λ According to Different Hypotheses: For Subjects (VaSs); for Stages (VaSt); For Subjects and Stages (VaSsSt); and, For Comparison, The Original Model With λ Constant For Subjects and Stages (CoSsSt).

	STAGE			
SUBJECT	VaSs (where $\lambda_{ij} = ci$)	VaSt (where $\lambda_{ij} = cj$)	VaSsSt (where λ_{ij} = random)	CoSsSt (where $\lambda_{ij} = c$)
	$\lambda_1\lambda_1\lambda_1\ldots$	$\lambda_1\lambda_2\lambda_3\ldots\lambda_K$	$\lambda_1\lambda_2\lambda_3\ \ldots$	$\lambda_1\lambda_1\lambda_1\ \ldots$
	$\lambda_2\lambda_2\lambda_2\ldots$	$\lambda_1\lambda_2\lambda_3\ldots\lambda_K$		$\lambda_1\lambda_1\lambda_1\ \ldots$
	$\lambda_N\lambda_N\lambda_N\ldots$	$\lambda_1\lambda_2\lambda_3\ldots\lambda_K$	$\ldots\ \lambda_{NK}$	$\lambda_1\lambda_1\lambda_1\ldots$

The effects of violating these assumptions were measured by (1) the mean of the sampling distribution of $\hat{k}$ and $\hat{\lambda}$ (2) the standard error of $\hat{k}$ and $\hat{\lambda}$, and (3) the goodness-of-fit by the appropriate gamma distribution fitted to the distribution of solution times of one experiment selected at random from the 200 that comprised each simulation study.

Since the value of λ is dimensionless and has no absolute meaning, any values could in fact have been used in the input arrays for the simulation experiments. The range of values over which λ was varied was given by

hypothesis, although this was suggested by results of the empirical studies (Chapter 4).

Two separate ranges of values were used, one more extreme than the other. A sample size of 45 was used throughout, and the results presented in Table 5.2 are the outcome of 200 repetitions of the experiment. In the VaSs paradigm, Range 1 varied from λ=.0990 to .2310, a ratio of approximately 1:2.4. The value of λ was increased in steps of size .0030 for each consecutive subject. The more extreme range, Range 2, varied from λ=.0800 to .3400, a ratio of approximately 1:4.3. λ was increased in steps of size .0600 for each subject.

Range 1 and Range 2 were tested separately over a series of K values from 1 to 7, where K=1,2,3,5 and 7. Table 5.2 shows the range of λ used when stages were allowed to vary, but when each subject had the same λ on a given stage (VaSt).

TABLE 5.2

Range of λ Used in VaSt Paradigm for Range 1 and 2, With Each Subject Represented By the Same λ Value on A Given Stage

K= 1	2	3	5	7
Range 1				
.1650	.1410	.1170	.1170	.0990
	.1860	.1650	.1410	.1170
		.2070	.1650	.1410
			.1860	.1650
			.2070	.1860
				.2070
				.2310
Range 2				
.2120	.1640	.1160	.1160	.0800
	.2540	.2120	.1640	.1160
		.2460	.2120	.1640
			.2540	.2120
			.2960	.2540
				.2960
				.3380

5.4 Results of Simulation Studies

In Table 5.3 are presented the results from the various manipulations of the size of λ. The behaviour of the original model with a constant λ value throughout, is shown for contrast.

TABLE 5.3

Means and Standard Errors (S.E.) of Sampling Distributions of $\hat{k}$ and $\hat{\lambda}$ For K=1-7 Under Three Experimental Conditions of λ Variation (VaSs; VaSt;VaSsSt) and Model Condition of Constant λ (CoSsSt)

		RANGE 1				RANGE 2			
K		VaSt	VASs	VaSsSt	CoSsSt	VaSt	VaSs	VaSsSt	CoSsSt
1	Mean k	1.1149	0.9870	0.9826	1.1053	1.0978	0.8470	0.8644	1.1053
	S.E.	0.2571	0.2827	0.2688	0.2812	0.2713	0.2366	0.2649	0.2812
	Mean $\hat{\lambda}$	0.1907	0.1562	0.1601	0.1129	0.2322	0.1578	0.1694	0.1129
	S.E.	0.0535	0.0551	0.0535	0.0389	0.0685	0.0582	0.0668	0.0389
2	Mean k	2.0453	1.8661	1.9334	2.1860	1.7843	1.4602	1.5164	2.1860
	S.E.	0.5073	0.5093	0.9751	0.4982	0.4504	0.4163	0.4598	0.4982
	Mean $\hat{\lambda}$	0.1541	0.1494	0.1522	0.1108	1.1502	0.1349	0.1448	0.1108
	S.E.	0.0425	0.0482	0.0427	0.0310	0.0469	0.1500	0.0541	0.0310
3	Mean k	3.0255	2.5610	2.8726	3.0857	2.7043	1.9135	2.3357	3.0857
	S.E.	0.8077	0.6549	0.7640	0.5391	0.5532	0.4622	0.5708	0.5391
	Mean $\hat{\lambda}$	0.1558	0.1352	0.1516	0.1033	0.1647	0.1160	0.1488	0.1033
	S.E.	0.0444	0.0415	0.0426	0.0250	0.0419	0.0356	0.0440	0.0250
5	Mean k	4.2523	3.7952	4.5822	5.4583	4.8268	2.5005	4.1126	5.4583
	S.E.	1.0605	0.8329	1.0881	1.1267	1.0955	0.5729	1.0362	1.1267
	Mean $\hat{\lambda}$	0.1520	0.1190	0.1436	0.1096	0.1826	0.0910	0.1533	0.1096
	S.E.	0.0347	0.0324	0.0362	0.0267	0.0479	0.0270	0.0445	0.0267
7	Mean k	6.8530	4.9264	6.4680	7.2495	5.8592	2.8114	5.1961	7.2495
	S.E.	1.6485	1.1874	1.5079	1.5174	1.4202	0.6872	1.1920	1.5174
	Mean $\hat{\lambda}$	0.1487	0.1101	0.1444	0.1041	0.1396	0.0728	0.1342	0.1041
	S.E.	0.0376	0.0290	0.0370	0.0237	0.0366	0.0206	0.0356	0.0237

The effect on the model will be described separately for the two ranges of λ that were used.

Range 1

The effects on the mean of the sampling distribution of $\hat{k}$, when λ is varied over the values of Range 1, is shown in Figure 5.1. The effect on mean $\hat{k}$, when K=1, is not appreciably different for different conditions, although the VaSs and VaSsSt conditions underestimate the population mean, while the VaSt and CoSsSt conditions overestimate it. The fact that the latter two conditions behave in the same way is not surprising since, when K=1, condition VaSt is identical to condition CoSsSt. The situation when K=2 and K=3 is much the same, although, from Table 5.3, we observe that when K is 3 or greater, the standard error of the three experimental conditions is larger than that of the condition originally described by the model (CoSsSt). The standard errors for the experimental conditions tend to be relatively stable at about one-quarter of the size of the value of the mean. When K=3, the condition VaSs deviates from, and underestimates the population mean by a larger absolute amount than do the other three conditions. (When K=3, $\hat{k}$(VaSs)=2.56.) This tendency increases as K increases, so that, when K=7, the mean $\hat{k}$ for this condition is 4.93. Condition VaSt consistently underestimates the true mean, and although the degree of underestimation increases as $\hat{k}$

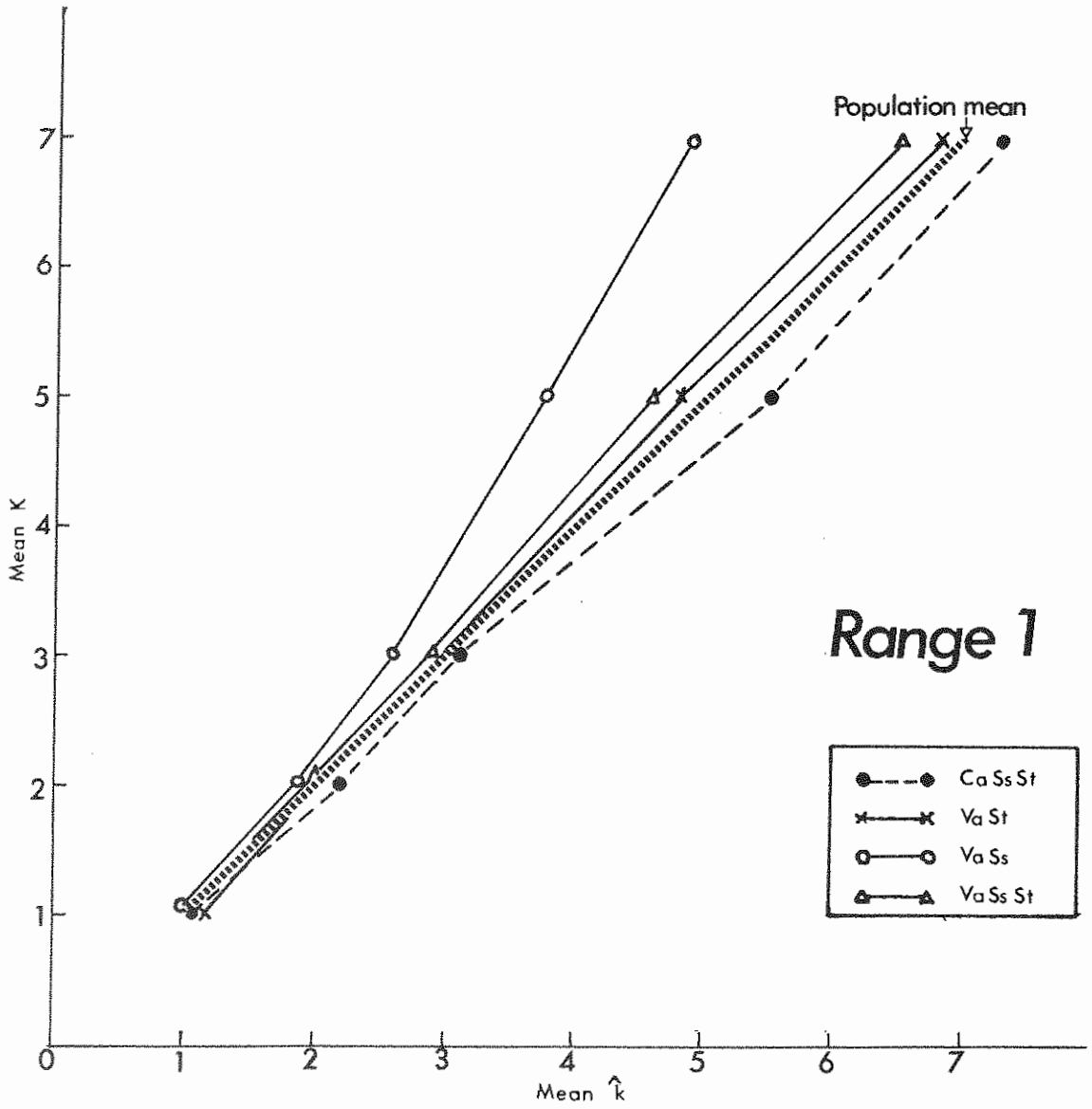


FIGURE 5.1: Mean of sampling distribution of $\hat{k}$ for K = 1 to 7 when λ varied over Range 1 values.

increases, it is never very great. As a result, Condition VaSt provides a less biassed estimate of the population mean than any of the other conditions, with a tendency to overestimate slightly when K is equal to or less than 3, and to underestimate when K is larger than 3. The situation described by the original version of the model consistently overestimates the population mean, with the tendency at a maximum when K=5, ($\hat{k}$=5.4583), and then decreasing again.

The total picture may be summarized by the statement that the situation in which the sample mean most closely estimates the population mean is the one in which all subjects are assumed to be alike in their propensity to solve a given module, but in which the group of subjects display different probability rates on different modules. This state of affairs more closely reflects the population mean than does the situation described in the model (CoSssSt). Violating the assumption of a constant λ produces a closer approximation to the 'true' situation than does the stages-module model stemming directly from the theory of waiting times.

Range 2

When a more extreme range of values is used, with a ratio of low to high values of 1:4.3 (compared to 1:2.5 in Range 1), the situation is essentially the same

as that appearing for Range 1, in that the ordering of conditions with respect to the degree of closeness to the population mean remains the same. However, the estimates provided by the three experimental conditions are not as accurate as in Range 1, even when K=1. This can be seen in Figure 5.2. In other words, the situation VaSs again presents the least veridical representation of the 'true' situation, consistently underestimating the population mean, but to a more extreme degree than was found in Range 1. When K=7, for example, $\hat{k}$(VaSs)=2.8114, and when K=5, $\hat{k}$(VaSs)=2.5005. The standard error (Table 5.3) remains constant relative to the mean. Condition VaSsSt is responsible for the next largest amount of bias, again in the direction of underestimating the population mean. This tendency increases as K increases, so that when K=7, $\hat{k}$(VaSsSt) =5.1961. Where λ is constant for subjects, but varied for stages, the situation very closely resembles that observed for Range 1, except that there is a sharp relative increase in the size of the deviation when K=5. Till this point however, the VaSt situation is the one that most closely reflects the population mean.

All three experimental trends change in direction when K=5, deflecting the estimated means further from the population mean when K is greater than 5. The opposite is

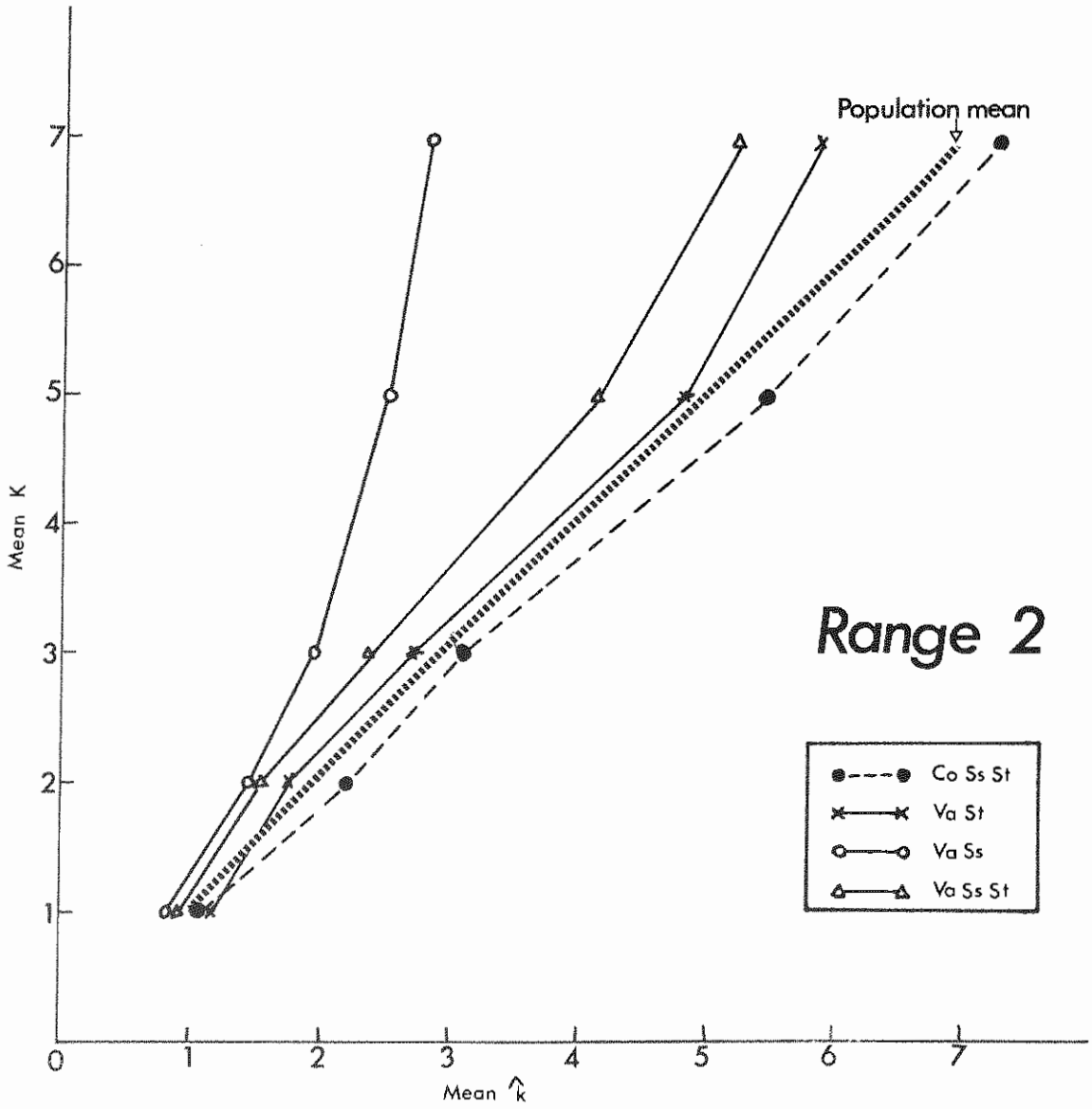


FIGURE 5.2: Mean of sampling distribution of $\hat{k}$ for K = 1 to 7 when λ is varied over Range 2 values.

true however when the assumptions of the model are not violated, so that the $\hat{k}$ values move closer to the population value when K=7. Till the point where K=5, the VaSt condition provides an exceptionally good estimate of the population parameter values. It is difficult to understand why the deviation should increase between K=3 and K=5 for all except the VaSt condition, since the actual range of λ used is the same for K=3 and K=5 (see Table 5.2), the only difference being that two intermediate values are added when K=5. This is true of both Range 1 and Range 2. When K=7 however, the ratio of lowest to highest λ value is larger than it is for any other K value, and this would explain an increase in the bias of the means of the sampling distributions of the three experimental conditions. It is interesting that the VaSt condition, which reflected the population values more closely in Range 1 than any of the other conditions, shows considerable bias in Range 2 only when K=7. Since constant values are used for both subjects and stages in the CoSsSt condition, it would be expected that the deviations from linearity would be constant irrespective of the value of K. The variations of this nature that do occur however,are probably largely due to the fact that only 200 trials constituted each experiment, and it is likely that with larger number of trials these deviations would disappear, although

the CoSssSt condition would still consistently overestimate the population mean.

In both Range 1 and Range 2, there is a high degree of consistency in the directions of the bias of the means of the sampling distributions of $\hat{k}$. The three experimental conditions (with the exception of condition VaSt for K=1 in Range 1 and 2 and, minimally, VaSt for K=2 and K=3 in Range 1) underestimate the population mean, whereas in the situation assumed by the model, the mean is consistently an overestimation of the true mean. The size of the standard errors relative to the mean appear to be stable throughout the various conditions.

The sampling distributions of $\hat{\lambda}$ is shown in Figure 5.3 and 5.4 for Range 1 and 2 respectively. A generalized description of the behaviour of these distributions is not possible. As would be expected, the CoSssSt condition produces a situation that is relatively stable, with little variance. This was inevitable, since the input for this condition consisted of λ constant at .1000. On the other hand, condition VaSssSt in Range 1 displays behaviour that is more invariant, irrespective of the size of K, than does condition CoSssSt. This suggests that randomly assigning different values of λ to subjects and stages has an effect on the model that is similar to the effect of a constant λ when the range of values assumed by λ is

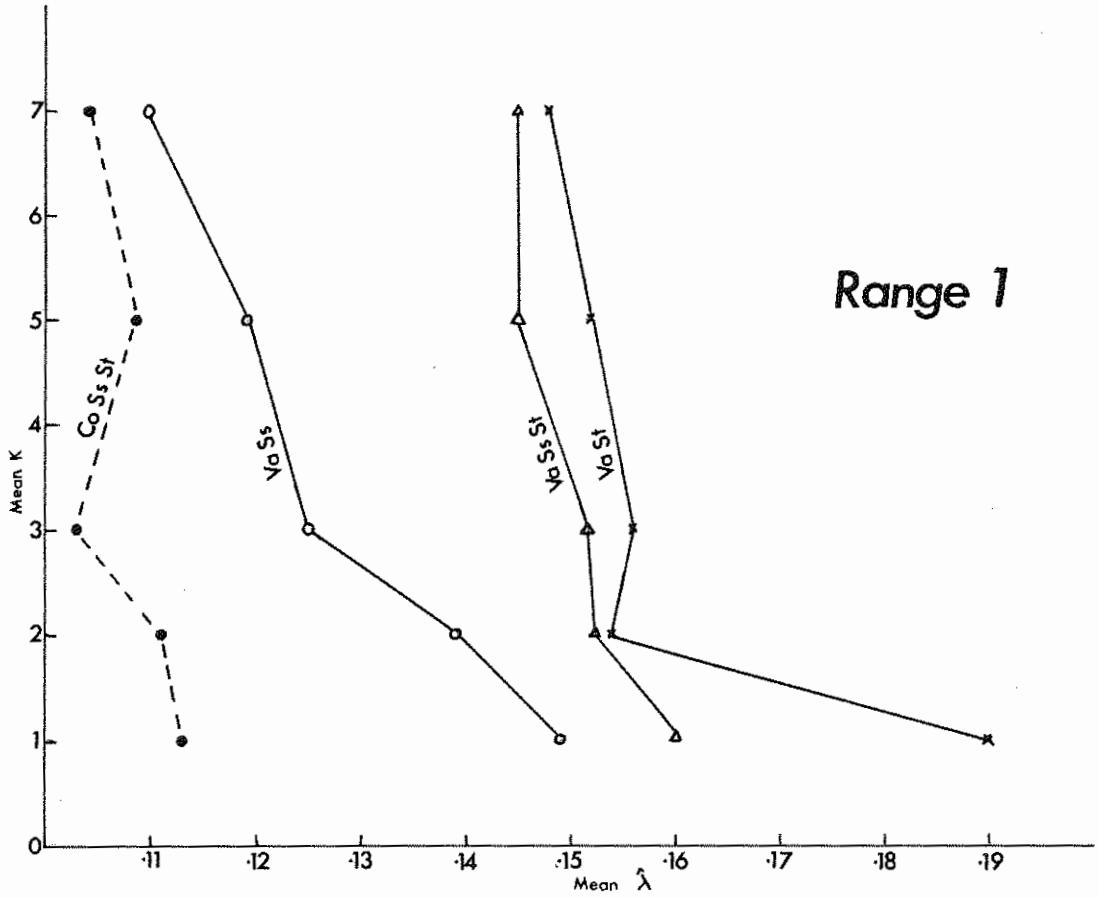


FIGURE 5.3: Mean of sampling distribution of λ for K = 1 to 7 when λ is varied over Range 1 values.

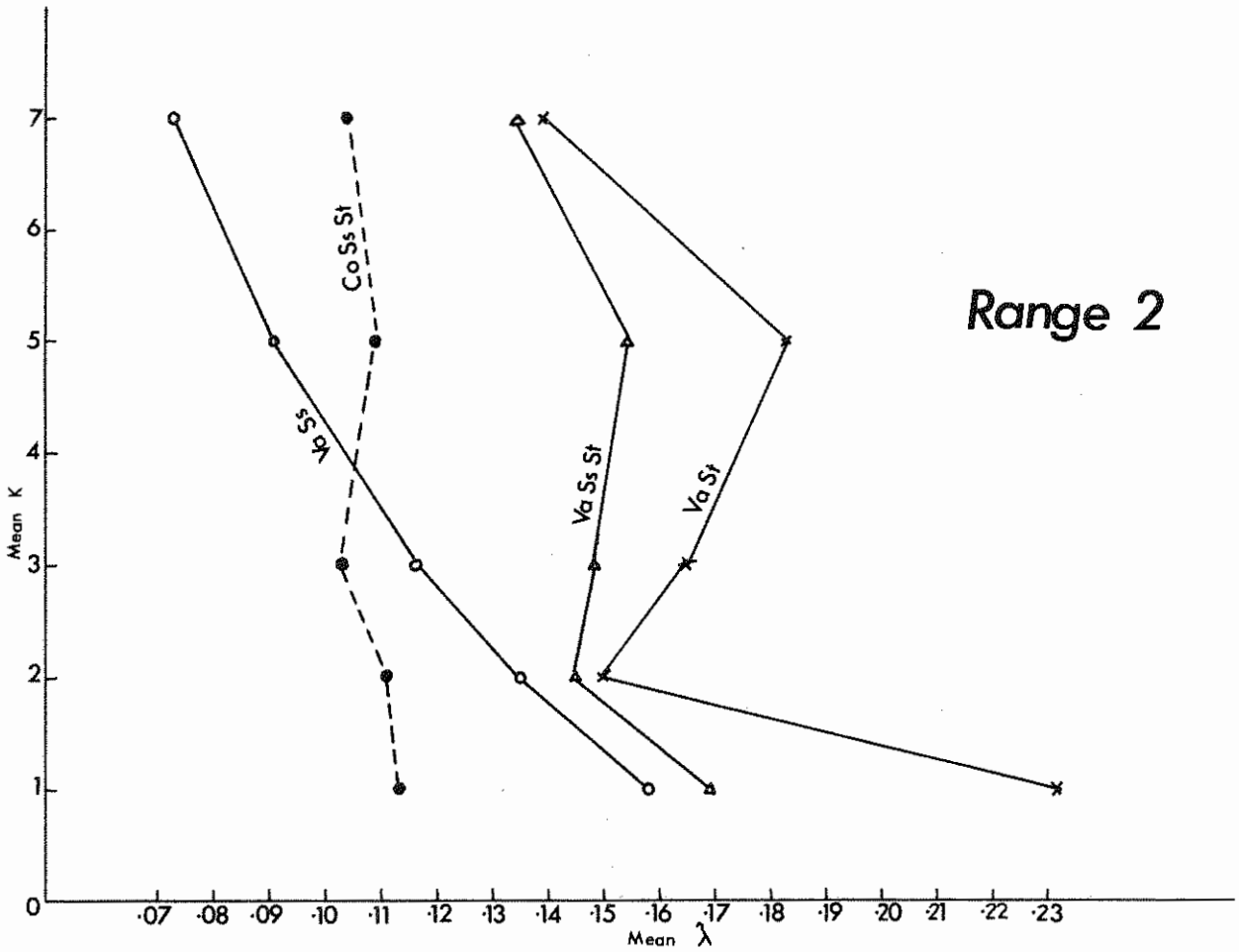


FIGURE 5.4: Mean of sampling distribution of $\hat{\lambda}$ for K = 1 to 7 when λ is varied over Range 2 values.

moderate. This situation is reversed in Range 2, however, so that condition VaSsSt shows more variability than CoSsSt, but the difference is not great. In contrast, when there is systemmatic variation of λ, for either stages or subjects, the means of the sampling distributions of $\hat{\lambda}$ show relatively large variations with differences in $\hat{k}$. Again, this is to be expected, since the larger the value of K, the larger the number of different λ values that were introduced into the experiment. It is not obvious why there should be linearity of relationship between mean $\hat{\lambda}$ and the size of $\hat{k}$ for the VaSs condition, since this would have been more appropriate in the VaSt condition, where the values are varied systemmatically for the various stages.

It might have been anticipated from the stages-module model that the value of $\hat{\lambda}$ would decrease with increasing K, since the more modules a problem has, the smaller the probability rate for its solution. Simply, the more complex a problem is, the harder should it be to solve. However this relationship appears clearly only in the VaSs condition, although it is suggested by the VaSt condition. This expectation should be independent of the conditions under which λ was varied so that a linear relationship might have been anticipated for the CoSsSt and VaSsSt conditions as well as for the others. Increasing the number of modules in a problem should increase the diffi-

culty of that problem, whether the modules themselves are similar or dissimilar in their difficulty levels. Further experimentation would be necessary to investigate what appears to be a shortcoming in the predictive adequacy of the stages-module model.

The final analysis of data obtained through Monte Carlo techniques concerns the cumulative distributions of solution times arising out of experiments in which various combinations of size of K, range of λ and experimental condition occurred. Appendix H lists these combinations, although the graphical representation of the cumulative distribution results are not included. These empirical distributions were fitted by the appropriate theoretical gamma distributions, and overall goodness-of-fit was evaluated for each distribution using the Kolmogorov-Smirnov one-sample test. It was found that gamma distributions, determined by appropriate values of $\hat{k}$ and $\hat{\lambda}$, do not, in any of the cases considered, differ significantly from these cumulative simulated distributions.

5.5 Conclusion

The simulation studies described in this section present us with the following results: although there is substantial variation in the degree to which different simulated conditions exhibit bias in estimating the population value of the parameter K, the cumulative

frequency distributions of solution times that comprise individual trials in the experiments underlying these sampling distributions are invariably fitted satisfactorily by the appropriate theoretical distribution. This remains true, irrespective of the size of the deviation as indicated by other analyses shown, for example, in graphical form in Figures 5.1 to 5.4. Since the same data is used in these two methods of testing the model, and since the sampling distributions do display, through the size of their mean and standard error values, sensitivity to the violations of the assumptions made by the model, it must be concluded that the gamma distribution is an inadequate model for the distribution of solution times generated in these experiments. It appears to provide an overly flexible fit to data which are known to violate properties assumed by the description. A model that is universally applicable will probably lack any meaningful explanatory value. Since the gamma distribution invariably describes the distributions obtained from both simulation and empirical studies (Chapter 4), care must be exercised in the interpretation of measures of goodness-of-fit. The point has been well expressed by Bush (1963) who writes:

> Having estimated the parameters of a model or theory, an investigator still wants to know how well his theory and data agree. He wants to measure 'goodness-of-fit'. This is sometimes viewed as a problem in hypothesis testing, but in my opinion this is a serious error. Model evaluation is not the 'acceptance-rejection' problem with which hypothesis testing deals. Factory products and men's hunches may be accepted or rejectd, but formal theories are evaluated modified, and extended. Thus we should not 'test' goodness-of-fit, we should 'measure' it. (p.432)

One way in which the gamma distribution could profitably be evaluated as a model would be by a comparison with the descriptive behaviour of some other theoretical distribution. Such a project, although it points beyond the scope of this thesis, represents an important step in future research.

CHAPTER 6

ANALYSIS OF THE PROBLEM STRUCTURE AND SOLUTION PATHS

Although the study of problem-solving behaviour forms a significant area of psychological research, little attention has been directed to the concept of a problem itself. We should pause to consider this concept in the present work since, as we saw in Chapter 3, it is essential for the experimenter to strive for an appropriate match between solver and problem if he is to have a useful vantage point from which to examine the problem-solving process.

Since, as we shall show, a problem cannot be defined without reference to the problem-solving system, it may be illusory to assume that an inclusive definition is possible.

Lying as they do on the same continuum, no absolute distinctions can be made between problems

and nonproblems, or between the trivial and the genuinely problemmatical. As has been said about the concept of a motive (Woodworth and Schlosberg, 1954), it may be that the term can never be made scientific, and that a potential problem must be viewed in terms of its interaction with a problem-solving system. This is not to say however, that the structure of a problem cannot be analyzed independently, as we shall see in later pages (Section 6.1). The distinction between a system with a problem and that problem is similar in spirit to the distinction between stages and modules (Chapter 2).

There are several ways of viewing a system that has been presented with a problem. Reitman (1965) defines a system as having a problem "when it has, or has been given a description of something, but does not yet have anything that satisfies that description" (p.126), which he then augments by the proposition that a problem is generated

> ...when we associate with a description of what we desire the requirement that an element be found, obtained or created that satisfies that description. (p.128)

A different version of a problem-solving system is provided by Newell, Shaw and Simon (1960) who consider that a problem exists when

> ... a problem solver desires some outcome or state of affairs that he does not immediately know how to attain. Imperfect knowledge about how to proceed is at the core of the genuinely problemmatic. (p.257)

These definitions explicitly assume that a problem-solving system knows it has a problem, which implies that some type of cognitive 'priming' must be included as an essential factor in any description of such a system. At this level, 'problem' becomes interchangeable with 'stimulus'. and the rules for problem solution may be subsumed under those that govern goal attainment. Goals and problems may be co-defined, since in both situations there is doubt about how to proceed from A to B, where B is a desired state. Whether this in turn implies that problem solving should be treated as a branch of learning theory however, is a matter that will be discussed in Chapter 8.

An alternative way of describing a problem-solving system is to consider that presented along with the problem, is a set of all possible solutions or paths to solution. This set may be very large, and the actual

solutions may be very widely and occasionally threaded through it (see Figures 6.1 and 6.2). Strictly speaking, the system is not given the solutions; rather, it has some process for generating the elements of that set of solutions. This generator may have algorithmic features, in which case it can be guaranteed that, should a solution to the problem exist, the generator will sooner or later produce it. If the generator has heuristic properties on the other hand, there is no guarantee that the correct solution will be generated, and therefore, that the problem will be solved.

Along with this capacity to generate and search a set of possible solutions, the system has also to be able to test that the solution has been reached. In other words, for each problem, if it is 'well-defined' (McCarthy, 1956), there must exist a systemmatic way of deciding whether a particular solution is acceptable. However, the search paradigm also includes situations where the subject does not know what he is looking for, although this is not a situation commonly encountered in psychological experiments.

In one sense, if an algorithm exists, the problem is trivial since solution is inevitable. However, in any interesting problem, the search through all possible solutions is too inefficient, which means that the essential quality

of search and consequently the generating of heuristic methods to make that search more efficient, is maintained. A problem is also trivial when there is only one alternative, or only one element in the set of solutions. A problem exists when there is a choice between several alternatives, and there is "imperfect knowledge about how to proceed" among them.

Mention of the features that characterize a trivial problem returns us to the distinction between problem and nonproblem, and consequently to the interaction between the problem solver and the problem. Since, as has been pointed out, there is no absolute distinction to be drawn between problem and nonproblem, the relative demarcations that do exist must be a function of the capacity and experience of the problem-solving system itself. In order to be able to distinguish between the responses of different systems, we must characterize the system itself. We may assume a continuum of response, so that what constitutes a problem for System A may be solved automatically for System B (in other words, for this system there is only one alternative in the solution set), and may be totally insoluble by System C.

This may be clarified by an example. Consider the addition problem "2 + 2 = ?". A five-year old child, presented with this problem, is in a state of 'imperfect

knowledge' about the solution, and continues in this state until he is able to supply the elements missing from the desired state or goal. On the other hand, the same problem does not exist as a problem for the normal adult since, for him, the solution is immediately available. (It is a matter for investigation whether the adult goes through the same procedures as the child, but simply at a highly accelerated speed, or whether some form of short-circuiting occurs to produce a qualitatively different response from that of the child). At the other end of system response might be that of a neonate for whom the problem does not exist as a problem either, but, obviously, in a very different way. For the neonate, the problem does not exist qua stimulus, while for the adult, the problem does not exist qua problem. As a general rule, it may be said that,given a state of attentiveness, or cognitive priming, a problem exists for system when it is impossible to guarantee that the capacity of the system is sufficient to produce the required solution. Although the words 'impossible' and 'guarantee' raise philosophical issues, these may be considered irrelevant for present purposes.

Even though the interaction of system and problem is critical to the definition of whether a problem

is genuinely problemmatical, it is possible to define the structure of a given problem in independent terms (just as one may describe the individual parameters of a problem solver independently of whether he is solving a particular problem). The structure of a problem exists whether or not anybody is attempting to solve it.

An analysis of the formal structure of a problem is useful as a basis for establishing problem taxonomies. It is also essential for understanding what a subject is trying to do when he is solving a particular problem, and for evaluating the relevance and appropriateness of the methods he employs. By understanding how an individual characterizes a problem, and into what general framework he places it, we can appreciate how his heuristics are ordered, and how they are transferred across problems according to his subjective classification system. Such a system would provide a means for validating an *a priori* classification based on the formal characteristics of a problem.

Given the need for a definition based on structure, suggestions are made for dimensions along which such a definition might develop. Reitman (1966, p.133) proposes that degree of specification of either or both problem statement (initial state) and required solution (terminal

state) be used. He gives as examples of two quite different degrees of specification the two problems of 'compose a fugue' and 'convert a sow's ear into a silk purse'. This index could be augmented or replaced by a vector or typology of methods required to solve a problem. Gross typologies are in existence already, such as 'addition' problems which require addition rules, but more subtle and generalized heuristic-based typologies are possible. Additional rules might include the classification of solutions according to whether they are unique, or self-confirming (Eureka problems),or purely idiosyncratic. The medium in which either problem statement or solution or both are expressed provides a further potential classification principle. Each of these proposed dimensions would in turn define the degree to which the solution and, therefore, the solver is constrained. An alternative way, and the one suggested by this research, would be to organize problems along a dimension of complexity indicated by the number of modules in a problem. However, since the two addition problems used in this study were estimated to have the same number of modules (Table 4.8), but since they clearly differ in degree of complexity (difficulty) as measured by mean solution times, (Table 4.1), the number of modules does not in itself

appear to provide sufficient discrimination between problems. This complexity index must therefore be supplemented by an ennumeration of the quantity and quality of elements within each module, or in the whole problem. Such a scheme would be limited to problems that are amenable to modular description, which would perhaps restrict its value.

In this chapter, the structural attributes of the three cryptoarithmetic problems will be analyzed to discover where the source of differential complexity lies (6.1). The analysis is made at a molecular level (although it does not go as deep as that made by Newell, 1967), to correspond to the level at which processing can be described in the protocols. This is followed by a description of the paths to solution taken by the subjects in the study, which is, in turn, built into the framework provided by the stages-module model. Addition and multiplication problems are treated separately (6.2 and 6.4). Suggestions are made about the organization of search patterns to complete a module, and about the source of most of the variation between individuals (6.5).

6.1 External Description

(a) Complexity analysis

The purpose of this analysis is to ennumerate the elements and situations that occur in each problem, and to provide a basis for comparison between problems. For example, the two addition problems each contain ten different letters, yet the arrangement of these letters produces two tasks that differ considerably in difficulty (see Table 4.1). In this section, an attempt will be made to explain the differences in processing times required for solving the problems by the differences in complexity of their internal structure. Table 6.1 contains an analysis of the letter elements of the three problems in which the number of occurrences of each letter is presented together with the column or columns in which it appears. The last row of the table contains the total number of elements in each problem.

The problems are presented below with the notation used to refer to a particular column:

Problem 1						Problem 2				Problem 3						
C1	2	3	4	5	6	C1	2	3	4	C1	2	3	4	5	6	7
D	O	N	A	L	D		S	I	R		H	A	N	N	A	H
G	E	R	A	L	D				I				M	U	S	T
R	O	B	E	R	T		P	E	E	P				S	T	U	D	Y
											T	U	E	S	D	A	Y

TABLE 6.1

Letter Analysis

Classification	Problem 1 (DONALD)	Problem 2 (SIR)	Problem 3 (HANNAH)
Number of rows	3	3	4
Number of columns	6	4	7
Number of different digits	10	5	10
Frequency of each letter	A:2(C4) B:1(C3) D:3(C1;C6;C6) E:2(C2;C4) G:1(C1) L:2(C5) N:1(C3) O:2(C2) R:3(C1;C3;C5 T:1(C6)	E:2(C2;C3) I:2(C3;C4) P:2(C1;C4) R:1(C4) S:1(C2)	A:3(C3;C6;C6) D:2(C5;C6) E:1(C3) H:2(C2;C7) N:2(C4;C5) M:1(C4) S:3(C3;C4;C6) T:3(C1;C4;C7) U:3(C2;C5;C5) Y:2(C7)
Total number of letter elements	18	8	22

This table shows the greater internal complexity of Problem 3 (HANNAH) in terms of more rows, columns and total number of elements, even though the number of letters that actually have to be decoded into digits is the same as for Problem 1 (DONALD). Problem 2 (SIR)

on the other hand, is clearly simpler than the other two, on any of the ennumerated attributes. In Table 6.2, the column is taken as the unit of analysis, and the problems are examined for the variety of column situations they possess. It will be seen from this breakdown that Problem 3 is again characterized by greater complexity than the other two problems, as shown by a larger range of column situations and by columns that are themselves more complicated.

TABLE 6.2

Column Analysis:
Incidence of Column Classes
by Column Identification

Column Class	Problem 1 (DONALD)	Problem 2 (SIR)	Problem 3 (HANNAH)
Columns with all letters different	2:(C1;C3)	4:(C1;C2;C3;C4)	4:(C1;C2;C3;C4)
Columns with two letters the same	3:(C4;C5;C6)	-	1:(C5)
Column with same letter in row and sum	1:(C2)	-	2:(C6;C7)
Columns with 4 rows	-	-	4:(C4;C5;C6;C7)
Columns with 3 rows	6:(C1;C2;C3;C4;C5;C6)	1:(C4)	1:(C3)
Columns with 2 rows	-	2:(C2;C3)	1:(C2)
Columns with 1 row	-	1:(C1)	1:(C1)

(b) Configuration analysis

The arrangement of letters within the column constitutes a further aspect of internal structure. The letters may all be different, or all the same or two letters may be repeated within a column. Clearly, these variations can be expected to lead to differences in amount of information that can be gained from the decoding of a single letter. We observe in Table 6.3 the configuration of distinct letters that appear in the three problems, followed by the number of cases of each in parentheses. The information presented in this table overlaps with the table of column analyses (6.2) but makes the column structure more explicit. Separate notation is used for each letter, so that the configuration (a+b+c=a) describes a four-letter column with three distinct letters, where the letter in the sum is equivalent to one of the letters in the addition. For this purpose, the situation where (a+b+c=a) is equivalent to (a+b+c=b) or (a+b+c=c).

TABLE 6.3

Configuration Analysis: Possible Letter Situations in Each of the Three Problems With Distinct Letters Referred to by Distinct Symbols

Problem 1 (DONALD)	Problem 2 (SIR)	Problem 3 (HANNAH)
a+a=b (3)	?=a(1)	?=a (1)
a+b=a (1)	axa=b (1)	a=b (1)
a+b=c (2)	axb=c (2)	a+b=c (1)
		a+b+c=d (1)
		a+b+b=c (1)
		a+b+c=a (2)

(c) Dynamic Processing Analysis

A third way of viewing the differences in problem structure is to consider the possible situations a solver is likely to confront during processing.[1] Let us assume we are observing the subject at any point during the experiment. His position in the path to solution may be located anywhere between the first exposure to the problem and successful solution. In other words, according to which problem he is working on, the solver could, on any given column, be in any of the information states contained in Table 6.4. In this table, an unassigned letter is denoted by l, a digit is denoted by d, and t

[1] This is closely related to the state model developed in Chapter 7.4.

represents a carry in and f a carry out of a column, which may or may not be known, so that each state has four levels, depending on whether one or both carries is known.

TABLE 6.4

Dynamic Processing Analysis: Possible Situations in Each of the Three Problems (where l=letter; d=digit; t=carry to a column, f=carry from a column)

Problem 1 (DONALD)	Problem 2 (SIR)	Problem 3 (HANNAH)
t+l+l=l+f	l=l	l=l
t+l+l=d+f	t+l=l+f	t+l=l+f
t+l+d=l+f	t+l=d+f	t+l=d+f
t+l+d=d+f	t+d=d+f	t+d=l+f
t+d+d=l+f	t+l=lxl+f	t+d=d+f
t+d+d=d+f	t+l=dxl+f	t+l+l=l+f
	t+d=dxl+f	t+l+l+d+f
	t+d=dxd+f	t+l+d=l+f
		t+l+d=d+f
		t+d+d=l+f
		t+d+d=d+f
		t+l+l+l=l+f
		t+l+l+l=d+f
		t+d+l+l=l+f
		t+d+d+l=l+f
		t+d+d+l=d+f
		t+d+d+d=l+f
		t+d+d+d=d+f

On this analytic dimension, Problems 1 and 2 are almost equivalent in the number of dynamic situations possible, and the total number of elements possible. In the light of empirical differences found between these two problems, this type of analysis seems to be less useful than the previous ones.

6.2 Path Analysis: Problems 1 and 3 (Addition)

Since the addition and multiplication problems demand separate treatment, this section will discuss the addition problems only, while the multiplication problem is treated in a later section. The module-stage estimates in the previous chapter indicate that subjects decoded the letters of the problem into their correct numerical representation according to certain invariant sequential patterns. In this section, the paths taken by the subjects will be examined in closer detail, which in turn will clarify the rationale employed in classifying the letter groupings as modules. 'Paths' refers to the sequence in which digits were correctly assigned to letters.

Consider first the situation where the problem is solved by purely exhaustive, algorithmic procedures, so that each letter is tested for every possible value. If there was no logical structure to the problem, the subject could have solved the problem in 10! different ways since there were ten different letters in the problem. In other words, there are ten possible entry points into the problem, and each of those ten letters could have been followed by the decoding of any of the nine remaining letters. (In Problem 1, however, since the numerical equivalent for D is given, there are 9! sequences in which the letters may be decoded.)

This is illustrated in Figures 6.1 and 6.2 in which a series of branching structures describe which, among all possible paths that a solver could theoretically follow, were the routes actually followed by the subjects. To avoid an excessively complicated diagram, all possible branches are not filled in.

The data for these results came from the protocols of the verbal subjects and from those nonverbal subjects who were able to reconstruct their solution paths, either on paper (Experiment II) or on tape (Experiment I). The data represent 27 out of 53 nonverbal subjects and 40 verbal subjects on Problem 1 (DONALD), and 20 out of 38 nonverbal and 29 verbal subjects on Problem 3 (HANNAH). Those who were unable to reconstruct their entire path sequence were excluded from analysis. The data for verbal and nonverbal subjects are pooled for each problem.

The paths followed are clearly a very small proportion of those that theoretically were available. A closer look at the figures reveals that, although subjects diverged at certain points, many converged again at others and remained on a common path to solution. In Figures 6.3 and 6.4, the data have been assembled into a form where common paths between individual letters can be seen, together with the number of subjects taking that route. The horizontal arrangement of the letters provides a description of which letters were

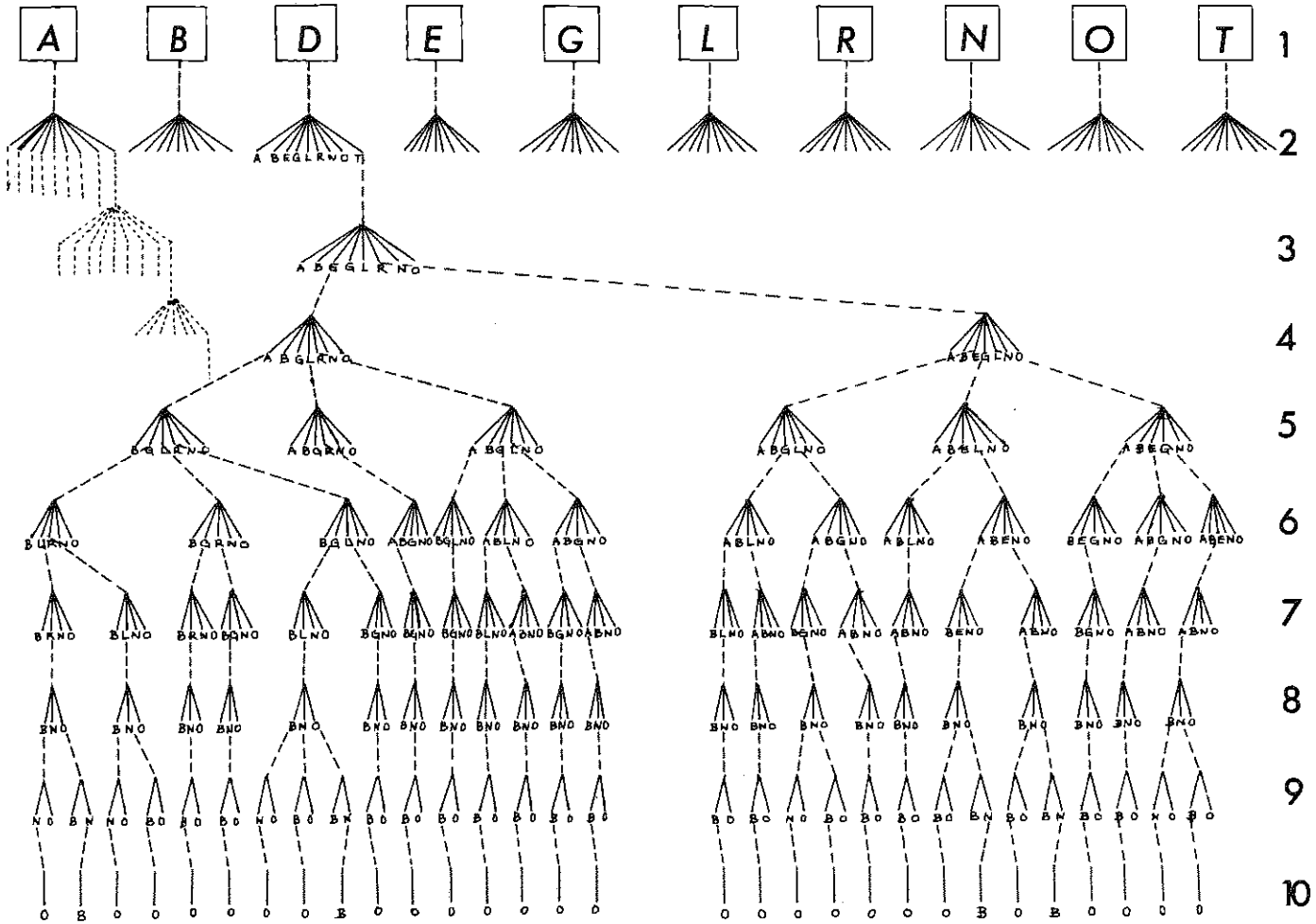


FIGURE 6.1: 'Tree' diagram showing which of 9! possible paths were generated by the subjects on Problem 1 (DONALD).

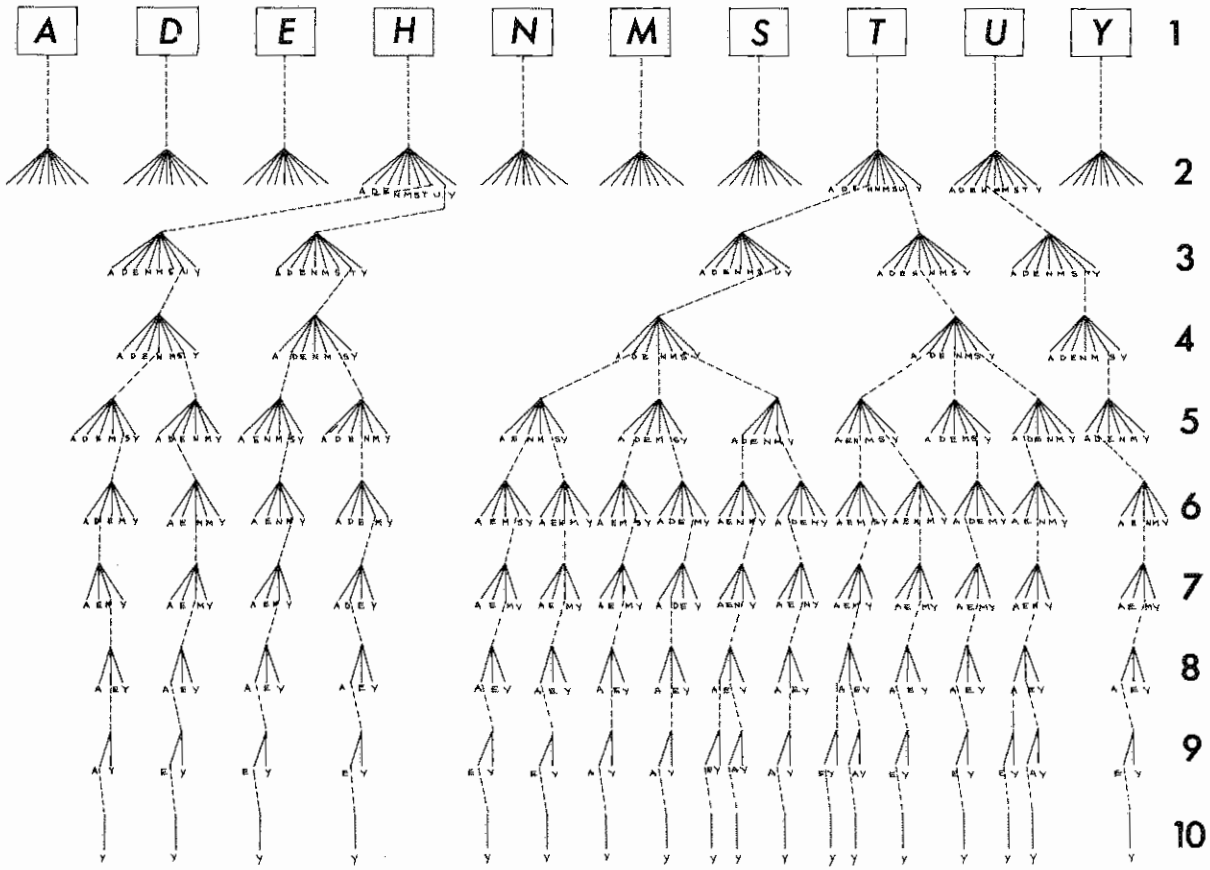


FIGURE 6.2: 'Tree' diagram showing which of 10! possible paths were generated by the subjects on Problem 3 (HANNAH).

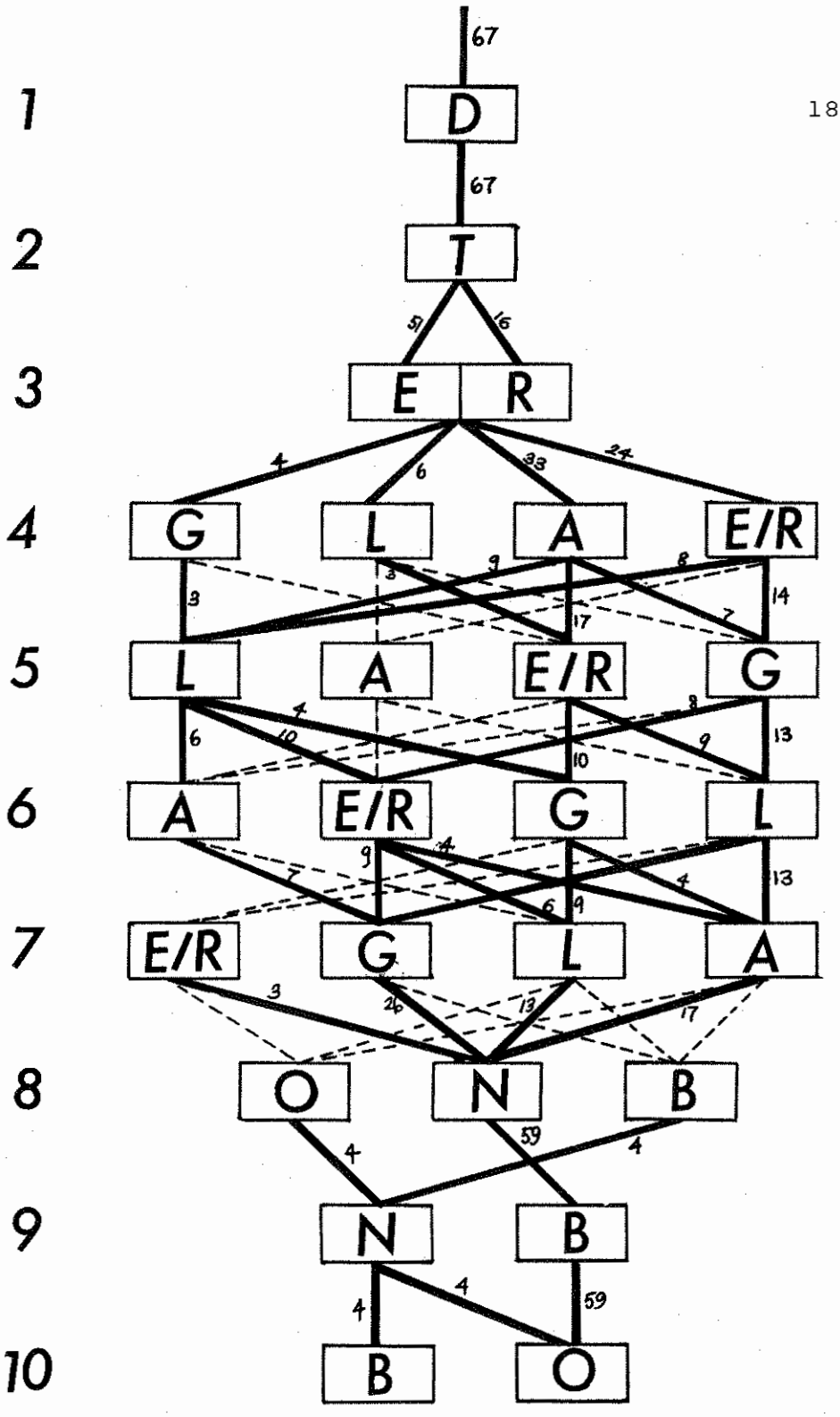


FIGURE 6.3: Problem 1 (DONALD). Order in which letters were decoded with the number of subjects following each inter-letter path. (Dotted lines indicate that path was followed by only one or two subjects).

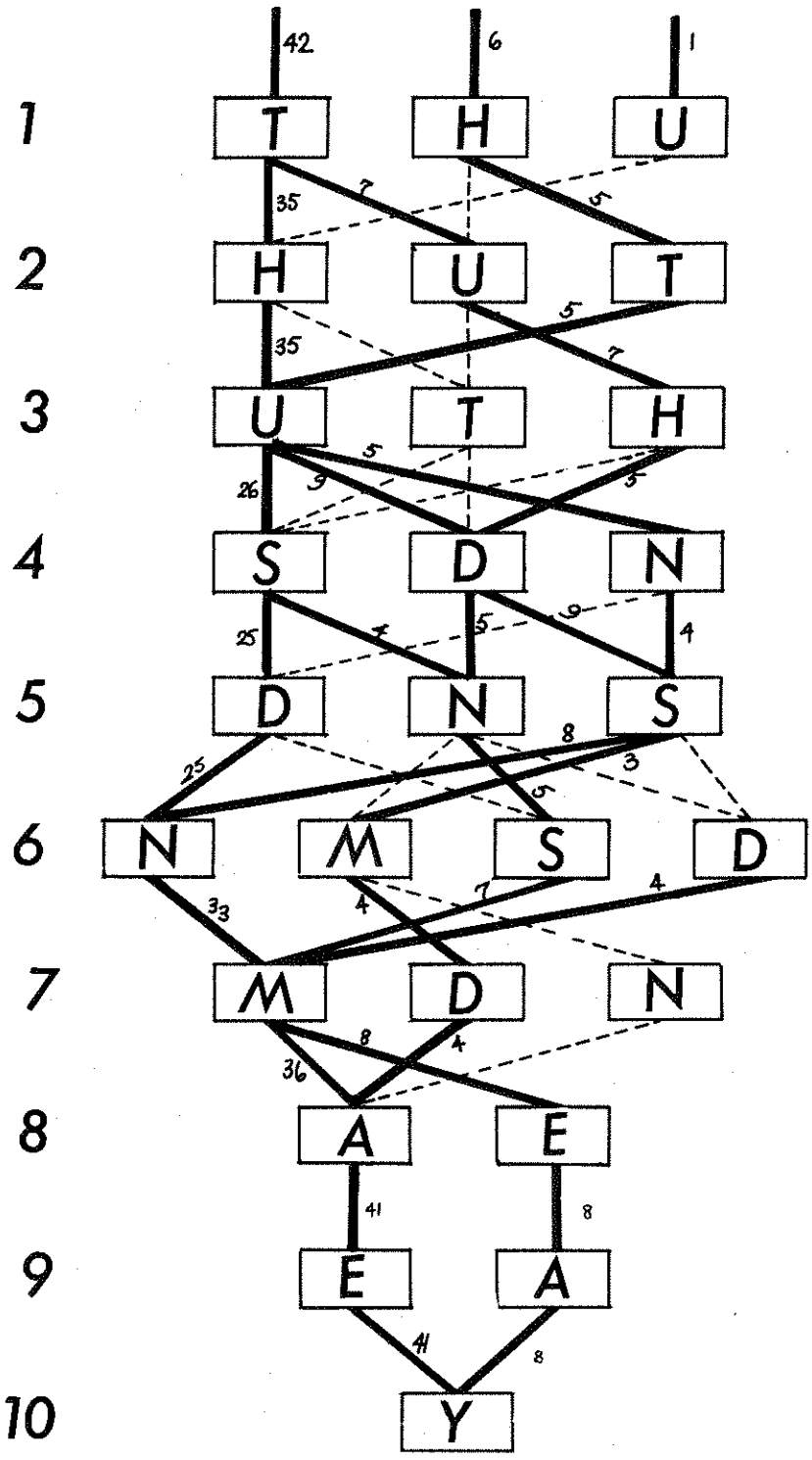


FIGURE 6.4: Problem 3 (HANNAH). Order in which letters were decoded, with the number of subjects following each inter-letter path. (Dotted lines indicate that path was followed by only one or two subjects.)

decoded at each subsolution point. For example, in Problem 3, the letters H,T and U were each the first to be decoded by some of the subjects, although 42 out of 49 decoded T first,and only one subject started with U. By contrast, because the numerical equivalent was given at the beginning of the problem, all subjects solved D first in Problem 1. The second letter to be decoded by all subjects was T, after which two paths, E and R appeared. E and R are treated as interchangeable since all subjects solved either E or R after T. The subsequent paths taken by either group are similar, with E or R substituted whenever appropriate. The numbers on the left-hand side of the figures refer to the number of correct assignments made up to and including that point, so that the letters G,L,A and E/R in the row marked 4, represent the fourth letters to be decoded. The dotted lines indicate that only one or two subjects followed that particular route. The major routes are represented by heavy lines.

The picture that emerges from these figures is rather diffused, especially in the middle section for both problems. Clearly, most of the variation in paths occurs during the substitution of the middle (4th to 7th) letters. However, the figures do substantiate the differences in difficulty between the two problems as displayed in the quantitative results (Table 4.1). There is, in Problem 1,

a greater choice of letters to decode at every point, while the higher degree of constraint in Problem 3 means that fewer paths were possible, so that the subjects had to spend more time searching for a correct entry point. Although the middle part of the problem involved the decoding of four letters in both problems (SDNM in Problem 3 and AGL(E/R) in Problem 1), levels 4 to 7 in Problem 1 have four letters that were decoded at each level, whereas in Problem 3, three of the four levels contained only three letters that could be decoded. The third part of each of the problems presented little difficulty to the subjects as can be seen from the short mean stage times for the last modules (see Table 6.7). It is not possible to compare directly the situations at the beginning of the two problems. The fact that the numerical equivalent for the letter D was already known, produced an artificially homogeneous path between the first two letters in Problem 1.

Scrutiny of Figures 6.3 and 6.4 will clarify the rationale involved in explicating the modules of the problems. The diffuse picture presented by these paths can be focussed into a clearer portrayal of the sequences followed by the subjects. Notice that the letters in the figures may be grouped into modules so that they form mutually exclusive groups which do not contain letters from any group. In other words, all subjects processed

certain groups of letters (i.e.a module) before proceeding, in fixed order, to process the next group, although the order of processing the letters within modules was variable. For example, in Problem 3, no subject decoded any letters in the SDNM module before letters in the THU module had been decoded. If this had occurred, one or all letters of the SDNM module would have appeared on the same row levels as those occupied by the letters T,H and U. This is not contradicted by the E/R grouping in Problem 1. Since this was a one-element module consisting of either E or R, whether a subject subsequently decoded E or R in the AGL(R/E) module was a function of which letter he had decoded at level 3.

The module theory becomes a useful way of describing the processing behaviour of subjects solving cryptoarithmetic problems. Figures 6.5 and 6.6 show letters regrouped into modules, together with the paths between modules and the number of subjects following them. Problem 1 is described as a 4-module and Problem 3 as a 3-module problem. The results described in Chapter 4 suggested that the middle module of Problem 3 is more accurately considered as two modules, since the $\hat{k}$ values were close to 2. Although this is a four-element module in the same way as Module 3 in Problem 1, it is much more complex. First, the range of possible values for each

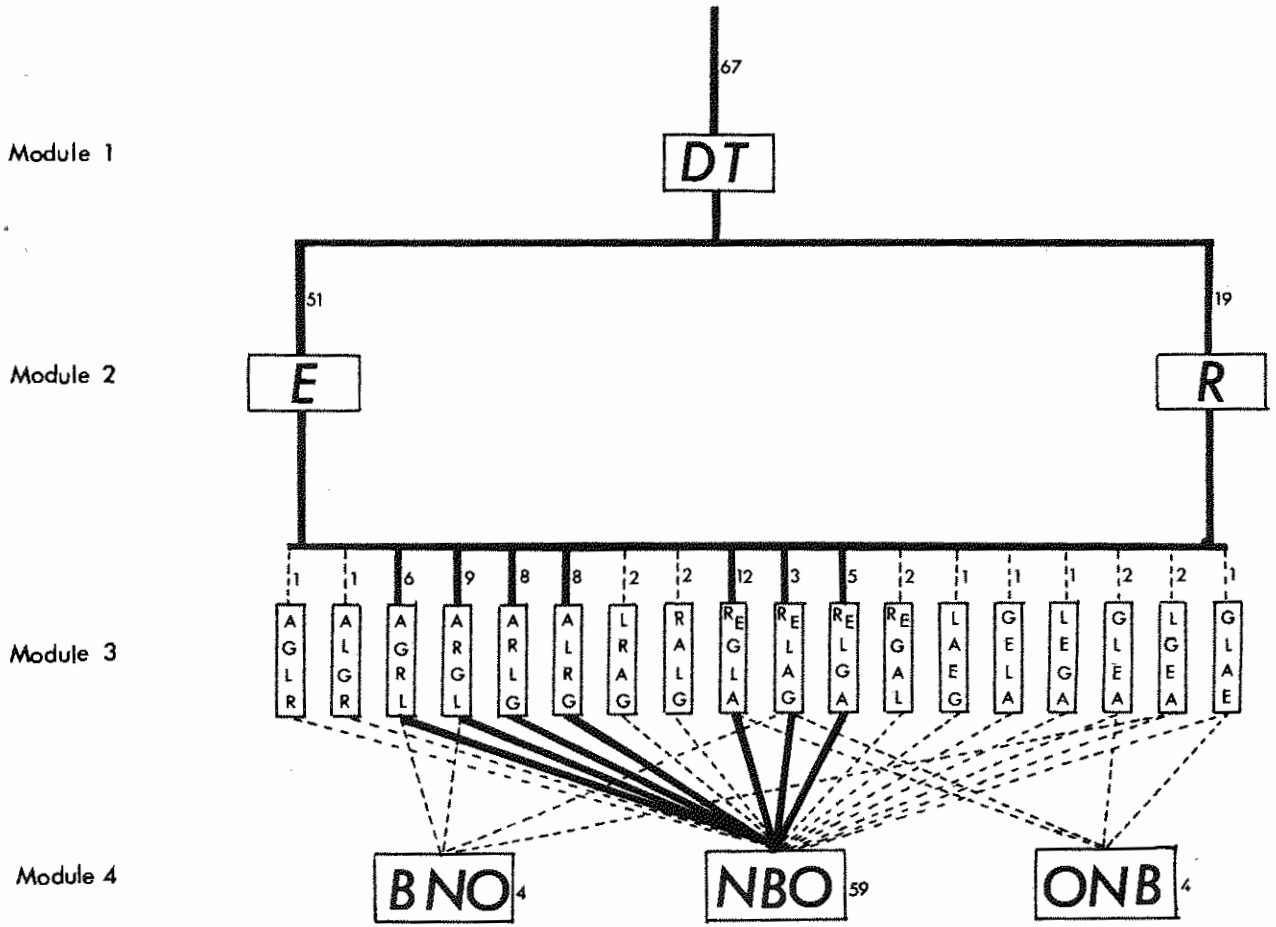


FIGURE 6.5: The stages-module model applied to the solution paths (N=67) of Problem 1 (DONALD)

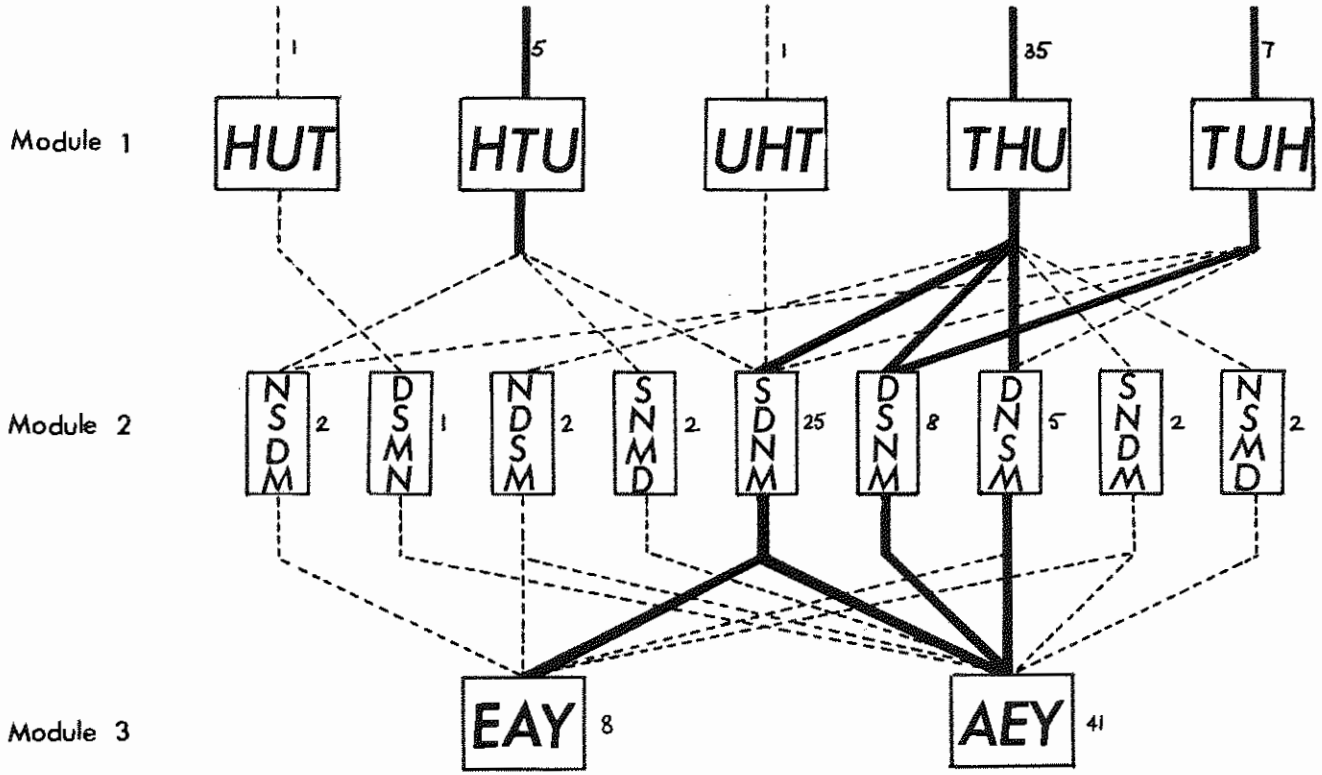


FIGURE 6.6: The stages-module model applied to the solution paths (N=49) of Problem 3 (HANNAH)

letter is larger, which means more time is required for ennumerating and testing the vectors of values. (This is discussed in greater detail in Chapter 7). Secondly, the four letters have a greater degree of inter-dependence, and their solution is governed by a larger number of ,and more complex equations than,the AGL(R/E) module in Problem 1. Once one digit is correctly assigned to a letter in the group, the others follow almost automatically, which means that most of the processing time in this module is search time, and only a small proportion of it is computational time. By contrast, the letters in Module 3 in Problem 1 have less internal dependence so that solution is less dictated by the decoding of any particular letter. Notice however that in all but four cases, M is the last letter to be solved. The within-module variability arises from the possible combinations of D,N and S. This will become clearer in Chapter 7, in which are presented artificial protocols, synthesized from the essential features of the subjects' protocols.

Although there are many arrangements of letters within modules, it would be erroneous to conclude that these are the results of the operation of a random process. Consider that although, in Module 3 of Problem 1, and Module 2 of Problem 3, there are 18 and 9

different arrangements respectively, it is possible to have 4! or 24 different arrangements in the absence of any heuristic or logical rules being applied to restrict the processing region. Notice again that Problem 3 is more constrained than Problem 1. It is possible that this difference would be removed if the assignment of the first letter had not been given but had to be made by the subject, although this is unlikely in the light of the structural analyses performed at the beginning of the chapter.

Figure 6.7 depicts the most common path between modules taken in the two problems, with the number of subjects taking it at each point. It is clear that the route from Module 2 to Module 3 in Problem 1 is the locus of most variation. The probability of deviating from the most trodden path is much lower in the more difficult problem, which suggests that the relatively high variance discussed in the preceding chapter arises from the time involved in searching for the few appropriate paths among a large number of alternatives. This matter will be resumed in more detail in the later pages of this chapter.

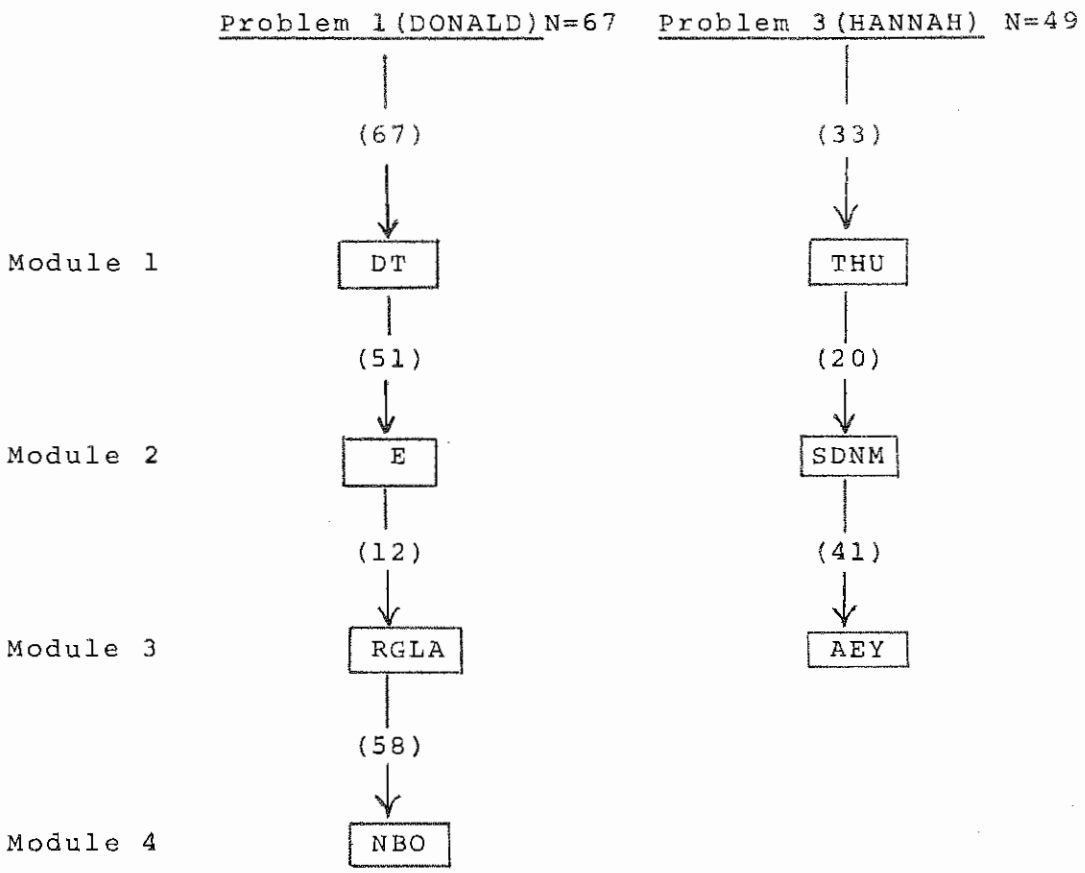


FIGURE 6.7: Most frequent inter-module paths with number of subjects taking each path ( ) for Problems 1 and 3

6.3 Transition Probabilities

In Table 6.5 and 6.6 are presented for Problem 1 and Problem 3 respectively, the observed probabilities that any two letters form an ordered pair in the decoding process, irrespective of where the pair occurs in the sequence of letter assignments. For example, in Table 6.6, whenever T appeared, the probability was .71 that the next letter to be assigned was H, whether T was the first or second letter in Module 1. On the other hand, the probability of T being followed immediately by Y is zero. Every letter is in an ordered pair relationship with at least one other letter.

The column transition probabilities for the middle and most difficult module of the problem indicate stricter relationships between ordered pairs in Problem 3 than in Problem 4. In other words, there are in Problem 3, fewer alternative assignments in each column and greater proportions of subjects contributing to a particular paired relationship. The first and last letters to be assigned are much more determinate in both problems, and in consequence, the relationships are more structured.

TABLE 6.5

Transition Probabilities Between Letters of Problem 1 (DONALD

	D	T	E	R	G	L	A	N	B	O
D	-	-	-	-	-	-	-	-	-	-
T	1.00	-	-	-	-	-	-	-	-	-
E	-	.76	-	.50	.05	.05	.25	-	-	-
R	-	.24	.31	-	.10	.16	.03	-	-	-
G	-	-	-	.25	-	.24	.21	-	-	-
L	-	-	.04	.25	.37	-	.19	-	-	-
A	-	-	.65	-	.09	.30	-	-	-	-
N	-	-	-	-	.39	.19	.25	-	-	.06
B	-	-	-	-	-	.03	.03	.94	.06	-
O	-	-	-	-	-	.03	.03	.06	.94	-
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

TABLE 6.6

Transition Probabilities Between Letters of Problem 3 (HANNAH)

	T	H	U	D	S	N	M	A	E	Y
T	-	.14	.04	-	-	-	-	-	-	-
H	.71	-	.16	-	-	-	-	-	-	-
U	.25	.71	-	-	-	-	-	-	-	-
D	.02	.10	.18	-	.55	.08	.08	-	-	-
S	.02	.04	.53	.22	-	.18	-	-	-	-
N	-	-	.08	.61	.25	-	.02	-	-	-
M	-	-	-	.08	.20	.71	-	-	-	-
A	-	-	-	.08	-	.02	.74	-	.18	-
E	-	-	-	-	-	-	.16	.82	-	-
Y	-	-	-	-	-	-	-	.18	.82	-
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

In Table 6.7 the stage values for individual modules, together with their mean times and variance are reproduced from Chapter 4 (Table 4.10) so that they may be reviewed in the light of increased qualitative information about the modules.

TABLE 6.7

Stage Values for Each Module and Mean Time (minutes) and Variance for each Stage for Problem 1 (DONALD) and Problem 3 (HANNAH)

Problem	Module Number	Letters	Stage ($\hat{k}$)	Mean Time	Variance
1 (DONALD)	1	DT	1.25	1.06	0.93
	2	E/R	1.44	8.75	7.55
	3	GLA(E/R)	1.22	12.76	12.24
	4	NBO	1.31	3.06	2.83
3 (HANNAH)	1	THU	1.12	8.05	7.89
	2	SDNM	1.72	22.62	15.86
	3	AEY	1.38	3.09	3.10

6.4 Path Analysis: Problem 2 (Multiplication)

The multiplication problem has certain characteristics that require it to be treated separately. It is distinguished from the addition problems since (1) it has four solutions, whereas the addition problems each have a unique solution, and (2) the product is a numerical palindrome and can therefore be solved by a 'trick', in this case realizing

that the product is divisible by 11, which in turn provides information about the structure of the multiplicand. A few subjects solved the problem by using this method.

More important, however, is the fact that (3) although there are clearly discernible 'chunks' in the processing of this problem, these differ in dimension from the ones described for the addition problems. The details of the problem-solving flow will become apparent in the next chapter, but a clarification of the key operations performed by the subjects is essential for an understanding of how the module model is to be applied to this problem.

In one sense, this is a one-module problem since all five letters are decoded in a single operation, following the correct assignment of one letter, the multiplier, (I). Examination of the protocols and the post hoc reconstructions of the nonverbal group suggests that there are three basic operational segments.

In the first, subjects looked at the problem as a unit, classifying it using equations and searching for an entry point. When this, usually the lengthiest of the operations, or group of operations, failed, the most typical response was to find ways in which the possible values for each letter could be limited. The result of

this operation was an accumulation of a series of limiting statements consisting of inequalities such as "I is not equal to 0, 1, 5" or statements like "R must be greater than or equal to 7". When further expansion of this series of operations was impossible, the subjects realized it was necessary to generate a subroutine to apply to the letters in the problem. This constituted the second operation, and the protocols contain many statements of the form "I'm tempted to set the thing up systemmatically" or "This is only a matter of trying various numbers. The rules aren't worth much". When this decision was made, nearly every subject then focussed on the letter I, although one or two used P instead, one of the letters in the product. I was set at different values, until a certain value led to no contradictions, with the column(IxR=P) serving as a unit for testing the generated values. When the(IxR=P)unit was satisfactorily assigned, the implications for the two remaining letters were tested, and, if no contradictions arose, the problem was solved. The third operational segment, the entry to the module itself, was considered to begin with the correct assignment for the value of I.

Because of the interdependence of all five letters in the problem, this is essentially a one-module problem. Since the problem was only given to the subjects in Experiment I, the small number of cases (16) did not justify analyzing

the protocols for stage times. What can be observed however, is that most of the time is spent on that part of the problem preceding the first correct assignment. The time from this point to solution is very short, with very little variation between subjects. This problem is distinguished by its relatively low variance (pooled verbal-nonverbal σ^2 =88.73). Presumably, most of the solution time and variance arises during the search preceding entry into the module. Some variance stems directly from whether the values for I are manipulated in ascending or descending order, and whether I has a high or low value for that particular solution. Much of the complexity of the solution process arises from the fact that the search itself consists of two phases as described above, but, by definition, this is not part of the module structure of the problem. The stage-module model is less satisfactory in describing this multiplication problem than it was for the addition problems, since it does not reflect the complexity inherent in the total solution process. Unlike the addition problems, where the complexity was distributed within and between modules, in this problem there is only one module which is not decoded until after the search phases have been completed. An additional inadequacy arises from the discrepancy between the pooled $\hat{k}$ estimate for this problem ($\hat{k}$=2.89) and the single module established through protocol analysis. In the absence

of information regarding the time to actually decode the module itself, it is impossible to ascertain whether this $\hat{k}$ estimate, with the time arising during the two search phases removed, would be close to 1.

An alternative way of integrating these results into the module framework would be to expand the concept of a module beyond that used here, by incorporating the infrastructure or superstructure of the problem. In the case of the multiplication problem, an expanded definition would include its palindromic property (superstructure), or its abstract composition in terms of a set of equations (infrastructure) so that SIR x I=PEEP can be abstracted to $(Rx10I+100S)I=IR+10I^2+100IS$. This augmented definition would enable the module structure to provide a more accurate reflection of the complexity of the processes that were used in solving the problem. Such a definition would not invalidate any of the previous analyses, but would simply make more explicit some of the infra- and superstructure characteristics inherent in the modules. For example, the decoding of the first three letters in Problem 3 implicity assumes operations that classify the problem as a three-row addition problem, and this classification in turn has implications for the size of the carries in and out of the columns of the problem.

It goes beyond the scope of this chapter to offer more than suggestions for the lines along which an improved module definition might develop. Clearly the present definition is not entirely satisfactory for the multiplication problem, mainly because it is a one-module problem. The most promising direction for future research would be towards the introduction of a concept that lies between that of stage and module to account for the qualitative aspect of the process. At the present, the theory describes only the time quality of the process and the structural properties of the problem, without accounting for the search that forms an integral part of the process, up to and including the processing of the elements in the module. Before developing the model further it would be necessary to collect protocols from a wider range of problems.

6.5 Search Operations

The previous section suggested that groupings of letter assignments are inadequate as the sole rationale for the delineation of modules in cryptoarithmetic problems. Clearly, the letter structure of the problem must be included in any description of the modules, but the adequacy of this framework for multi-module problems decreases when there is only a single module. It has been proposed that this deterioration occurs because the definition ignores the search operations leading up to the

decoding of the first letter of the module. In a multi-module problem, the search characteristics, both temporal and operational, are distributed between and within several modules. The protocols indicate that it is misleading to over-emphasize the actual assignments of digits to letters. Problem solving does not consist of a constant flow along one continuum, but is more accurately represented by holding operations during the search phase or phases where methods are searched, equations manipulated, etc.,and, with a change of state, output operations during the digit assignment phases at each module. The problem solver encounters the module, as we have defined it in terms of correct letter-digit substitutions, only at the end of a stream of processing information. Most of the stage time spent on a problem is pre-module search time and most of the operations performed to complete a module occur before the first element of the module is correctly assigned. Once entry into a module has occurred, digit assignments are made rapidly. In other words, the stage time can be divided into between- and within-module time with most of the time building up between modules, while most of the variance between subjects is the variance of inter-module search.

In Table 6.8 and 6.9 the times for the verbal group that were first presented in Table 4.12 are recalculated to extract the mean times and variance between and within

modules. For these calculations, the search time to the first module consisted of the time from when the subject first started his stopwatch to the correct assignment of the first letter in Module 1. The actual module processing time consisted of the time from the first to the last letter-digit substitution in that module. This procedure was repeated for the other two modules. An exception arose in the case of Module 2 in Problem 1, which consisted of only one letter so that no within-processing time can be calculated. (The sum of the mean times in the following tables are not equal to,but less than the total mean times of Table 4.1 since some time usually elapsed between the last correct assignment of the problem and the subject's report of 'I have finished'.)

TABLE 6.8

Mean Stage Times (minutes) Analyzed into Between- and Within-Module Time: Problem 1 (DONALD) N=40

	Mean Time	Variance
Beginning to Module 1	0.63	0.50
Within Module 1	0.43	0.39
Between Module 1 & 2	8.75	71.64
Between Module 2 & 3	5.55	62.42
Within Module 3	7.21	34.34
Between Module 3 & 4	2.58	7.74
Within Module 4	0.48	0.34

In every case, the variance between modules is larger than the variance within modules. The same is true of the mean times except for Module 3 where the mean time spent within the module exceeds the time preceded by search. It is difficult to explain this, although some of the explanation might lie in the fact that Module 3 is a module with a relatively high degree of independence between its elements, which implies that time spent searching between letters is longer than when letters are highly interdependent. An alternative explanation might be that Module 3 is not a unitary module, although this is not likely in view of the results for Module 2 of Problem 3 below.

The between-within variance hypothesis is completely supported by the results of Problem 3 in Table 6.9.

TABLE 6.9

Mean Stage time (minutes) Analyzed into Between and Within Module Time: Problem 3 (HANNAH) N=29

	Mean time	Variance
Beginning to Module 1	4.62	28.10
Within Module 1	3.43	8.08
Between Module 1 & 2	21.18	222.66
Within Module 2	1.44	9.45
Between Module 2 & 3	2.15	5.83
Within Module 3	0.94	0.54

These results lead to the hypothesis that if search time were controlled, the mean and variance of total solution time would be reduced. The protocols suggest that the mean would be reduced more than the variance, since the effect of controlling the search time would be to decrease the time for each subject by a certain amount. It is possible however, that the amount of reduction would be constant, in which case both mean and variance would be affected. Alternatively, this procedure might affect each subject by a constant proportion, so that the main effect would be on mean time, leaving variance relatively unaffected. These hypotheses suggest a future experiment in which the problems would be presented to the subjects along with an outline of the modules. For example, the subjects would be told that certain groups of letters were to be processed before others, without necessarily any suggestion of within-module letter order being included. In such a situation, it is likely that much of the search time would be eliminated, with the implication that this would result in a closer correspondence between the $\hat{k}$ estimate and the 'real' K, by allowing the natural structure of the problem to emerge uncontaminated. An experiment developed along these lines would provide a direct test of the stage-module theory by confirming or rejecting the prediction that the value of $\hat{k}$ estimated from the solution times data would equal

the number of modules presented to the subjects.

To control and reduce the search time is not to deny the significance of the search phases during problem solving. Although the search phases are a necessary component of any problem-solving process, they also contain many random elements and individual differences which may be responsible for the failure of the model to provide a reliable estimate of the number of modules (Chapter 4). By performing this experiment, the differences between verbal and nonverbal conditions would be clarified, although it is not possible to predict the direction of these effects independently of the level of problems that would be used.

6.6 Conclusion

We have seen that the structure of complex problems may be analysed in terms of its constituent elements, which suggests a classification scheme for cryptoarithmetic problems, if not one that may perhaps be extended to other types of problems. Given that performance indices can, at least in part, be traced to these elemental factors, the structure of the problems was described in terms of solution paths followed by the subjects in the study. Bounds are imposed by the use of heuristic methods which will be investigated in detail in the next chapter, which drastically reduce the size of the region within which search proceeds. Again, problem complexity was manifested in the number of

solution paths that are followed: it would appear that an easier problem can be distinguished by being associated with a larger subset of correct alternatives in the total set of alternatives through which a problem solver must search. Another way of expressing the same data is by transforming solution paths into matrices of transition probabilities between letters.

Although an extreme interpretation of the stages-module model would admit only one universally followed path series, the actual application of the model provided an excellent description of a restructured version of the data (Figures 6.5 and 6.6), an encouraging validation of the qualitative (module) aspect of the model. However, the shortcomings of the model were evident when the solution process for the multiplication problem were examined. Suggestions were made for the development of a construct to handle the search aspects of the solution process as a supplement to the purely temporal-structural status of the model as it now exists. The stage times to solution of a module, when analyzed into between- and within-module time, pinpointed some of the variability of the solution times data.

This discussion concludes with the contention that the module structure of a problem suggests a valuable framework for distinguishing between phases of the search process, which, in turn, requires future experimentation to further examine the behaviour of the stages-module model.

CHAPTER 7

ANALYSIS OF PROTOCOLS

> If people must be this complicated, and if things this complicated cannot be studied experimentally, then scientific psychology must be impossible. (This)... is a complaint that, if true, would certainly be important to prove.
>
> But is it true? Certainly, if one interprets 'scientific' to mean that all of a subject's verbal responses must be ignored, then it will be impossible to study thinking at the level of complexity required for programming computers or for understanding the neurology and physiology of the brain. But are such Spartan strictures necessary? They would protect us from long, violent disputes about 'imageless thought' perhaps, because they would make it impossible to say anything at all about thoughts, but that is a high price to pay for consonance. (Miller, Galanter & Pribram, Plans and the Structure of Behaviour, 1960, p.192)

Problem solving is more than the chaining together of a series of discrete steps. Any comprehensive account of a problem-solving attempt must include the various levels and types of activities performed by the solver, and so permit distinctions to be made between individuals and situations. These distinctions lie concealed in the quantitative and structural analyses as exemplified in earlier chapters, but they can be excavated and made visible through the application of information-processing techniques to verbal protocols collected in studies such as

this one.

An information-processing theory is concerned with the fine structure of behaviour, and postulates a detailed and precise set of mechanisms - functions and processes - to account for the observed behaviour. An information-processing system is dynamic - it engages in active, searching, analytic behaviour; it is purposive, manipulating information and objects in its path towards some goal. Such a view contrasts sharply with that of an organism, passively positioned in some multidimensional space, which is defined in terms of variables deduced statistically from large populations. Research guided by this latter school of thought is aimed at specifying the variables and factors that underlie behaviour. Information-processing theory, on the other hand, has as its objective the construction of functioning models of psychological structures and processes that will reproduce, and therefore explain, behaviour. The difficulty in analysing the actions of a problem solver as an information-processing system lies not in the intractability of methods or data, but 'in an embarrassment of riches'. It is hoped, however, that the outcome of such analysis will yield a small number of elementary processes which, when properly organized into some integrated activity, will provide a sufficient explanation of the solution process for a certain problem.

The first section (7.1) of this chapter contains 'ideal' protocols, synthetic reconstructions of salient activities observed in each of the three problems. This is followed by a discussion of the effects and mechanics of the verbalization process (7.2). The third section provides a detailed report of the language used by Newell (1966) to describe the behaviour of a subject solving a cryptoarithmetic problem, and this is illustrated with examples from protocols of subjects in the present study (7.3). The fourth section (7.4) attempts to evaluate the stages-module model in the context of information-processing theory, and is followed by a fifth section (7.5) in which a supplementary construct, the state model, is developed. The final section (7.6) of this chapter has as its objective the construction of a classificatory scheme by which to describe the major mechanisms that recur in the behaviour of the subjects on all problems contained in the study.

7.1 'Ideal' Protocols

A problem solver is viewed here as having constructed a search model (Johnson, 1944), which is an internal representation of the requirements of the solution as understood by him. Faulty construction of the search model will lead to errors, and random behaviour will occur when no model can be constructed.

Constructing a search model for cryptoarithmetic problems necessitates generating both heuristic and algorithmic methods. As will become clear in the examples that follow, there are points in the solution process that demand the application of an exhaustive trial and error routine; this routine does not appear in haphazard fashion, nor does it represent a breakdown of reasoning of logic. The most prominent use of trial and error appears in the multiplication problem and in the processing that leads up to the solution of the last module of the two addition problems. Although heuristic methods are used to impose limits on the region within which search occurs, that is, precisely which set of digits will be generated, the fact that all digits are then tested, and all possible paths generated within this region, means that the method used at the point is an algorithm. Given that the region has been correctly bounded, then the algorithm guarantees success. The algorithm routine consists of two mechanisms: (1) a generator that produces new combinations of values, and (2) a verifier that determines whether each new combination is in fact a solution to the problem.

A thorough examination of the protocols produced in the experiments led to the construction of a hypothetical protocol from an 'ideal' problem solver on each problem. Each hypothetical protocol represents a synthesis of correctly

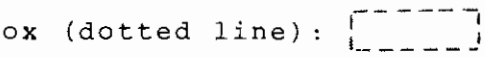
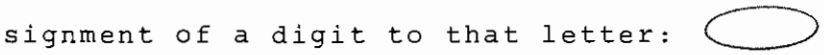
applied principles, and portrays the most efficient use of heuristics without the errors, backtracking, forgetting etc., that characterize a genuine subject.[1] This exercise is valuable in that it creates a context within which to view the solution process, and from which to evaluate deviations and idiosyncracies displayed by the subjects. In a sense, it parallels the use of a prototypical learning curve.

Turning now to the 'ideal' protocols themselves, some comments are in order to explain the notation that is used. The arrows between protocol statements represent the flow of processing, but have no quantitative meaning: the number of arrows in a protocol is not necessarily an indicator of the relative length of time for processing that problem. The following symbols are used to express the actions that occur in the 'ideal' protocols:

- a box (solid line) around an equation represents a correct assignment of a digit to a letter: 
- The position of letters and digits in the problem are referred to by a column/row notation, so that C1R3 refers to the letter/digit in Column 1, Row 3.

---

[1] Nine protocols are reproduced in full in Appendix I. They were selected to represent a wide range of behaviour, and are accompanied by a commentary detailing the salient features of each protocol.

- A search decision, instruction or question is expressed in capital letters, and the result of such processing is indicated by ⟶ :
- The summarizing of known information is displayed in a box (dotted line): 
- Parentheses occur when an implication is made, or an inference, reason or argument for an action is expressed: ( )
- $\begin{matrix}\cdot\cdot\\\cdot\cdot\end{matrix}$ indicates the result of testing a letter or a set of letters by an algorithmic routine.
- An oval around a statement indicates that there has been an increase in the specificity of information for a particular letter or letters. This indicates a situation somewhere between no information for a letter and final assignment of a digit to that letter: 
- (-p) indicates that the outcome of an argument is not possible; (p) indicates that the outcome is possible.

This representation of the solution process combines paths taken with their associated reasoning processes. It is not, strictly speaking, a flowchart, and it is for this reason that conventional information-processing notation has not been used.

The notation is not dictated by any theoretical considerations, but simply is intended to provide a convenient shorthand for representing various dimensions of

the solution process. The intention is to distinguish between those aspects of the protocol that are concerned with search procedures, subgoal attainment (digit assignment), access to previously established information, inferential and implicational reasoning, computational procedures, the products of search, as well as to present an overview of the main features of each problem. The categories are mutually exclusive though not exhaustive.

The protocols from the two addition problems will be presented first, and their similarities and differences discussed. The multiplication problem protocol follows on this discussion.

(a) 'Ideal' Protocol for Problem 1

	C1	2	3	4	5	6
R1	D	Ø	N	A	L	D
R2	G	E	R	A	L	D
R3	R	Ø	B	E	R	T

1.1 Given [D=5] —>
2 FIND D —>: there are 3 D's: C1R1;C6R1;C6R2 —>
3 PROCESS C6: 5+5=T —> : [T=0] —>carry 1 to C5 —>
4 FIND A COLUMN TO PROCESS —>:
5 DO A DISPLAY —>
6 [5ØNAL5]
[GERAL5]
[RØBERO] —>
7 FIND A COLUMN TO PROCESS —>:
8 PROCESS NEXT COLUMN: C5: L+L+1=R —>: R odd —>
9 FIND AN R —>: there are two R's:C3R2;C1R3 —>
10 PROCESS C3: N+R=B —>: 2 unknowns, no processing possible—>
11 PROCESS C1:5+G=R—>:R>5(no carry from C1,R≤9)—>: (R=7,9)—>
12 FIND AN R —>: no more R's —>
13 FIND A COLUMN TO PROCESS —>:

```
14 C1,C3,C5,C6 already scanned ——>
15 PROCESS C2——>:Ø+E=Ø——>:E=0,9(but T=0)——>:[E=9]——>carry 1 to C1,
16 (Ø=free) ——>                                    carry 1 to C2——>
17 ANY OTHER E's?——>:C4R3:A+A+1=9 ——>
18 DO DISPLAY ——>:
19  5ØNAL5
    G9RAL5
    RØB9RO ——>
20 FIND A COLUMN TO PROCESS ——>:
21 PROCESS NEXT COLUMN:——>C4:A+A=1=9,19——>:(A=4,9) (but E=9)→:[A=4
22 FIND A COLUMN TO PROCESS ——>:
23 ASSESS DEGREE OF COLUMN ASSIGNMENT——>:  fully assigned:C4,C6
                                           part assigned:C1,C2
                                           unassigned:  C3,C5
24 WHICH LETTER OR COLUMN CARRIES MOST INFORMATION?——>:
   R occurs three times: C1R3; C3R2;C5R3——>
25 WHAT IS KNOWN ABOUT R?——>:R=7,9(but E=9)——>: [R=7] ——>
26 FIND A COLUMN TO PROCESS——>
27 ASSESS DEGREE OF COLUMN ASSIGNMENTS——>:  fully assigned:C4,C6
                                            part assigned: C1,C2,
                                                           C5
28 C2,C3: not enough information——>
29 PROCESS C1,C5——>
30 PROCESS C5 : L+L+1=7,17——>(L=3,8)(but carry 1 to C4)——> [L=8] ——>
31 PROCESS C1:5+G=7——> (G=1,2) (but carry 1 to C1)——>[G=1] ——>
32 FIND A COLUMN TO PROCESS——>:
33 DO DISPLAY——>:
34 5ØN485
   197485
   7ØB970
35 LIST UNASSIGNED LETTERS——>:Ø,N,B——>
36 LIST UNASSIGNED DIGITS——>:2,3,6——>
37 LIST PART ASSIGNED COLUMNS——>:C2,C3——>
38 FIND A COLUMN TO PROCESS——>:C2:Ø+9+1=Ø——>:(Ø=free, no
     processing possible)——>
39 FIND A COLUMN TO PROCESS——>: C3: N+7=B——>:
40 USE TRIAL AND ERROR WITH REMAINING DIGITS——>
41 SET N=2,3,6 IN ASCENDING ORDER——>:
42 SET N=2——>2+7=6(-p)——>
              2+7=3(-p)——>:: N≠2——>
43 SET N=3——>3+7=2(-p)——>
              3+7=6(-p)——>:: N≠3——>
44 SET N=6——>6+7=2(-p)——>
              6+7=3 (p)——>:: [N=6] ——> [B=3] ——>
45 FIND A COLUMN TO PROCESS——>:C2:Ø+9+1=Ø——>
46 LIST UNASSIGNED DIGITS——>:2——> ASSIGN [Ø=2]
47 DO DISPLAY——>:
```

48 $\boxed{\begin{array}{r} 526485 \\ 197485 \\ \hline 723970 \end{array}}^{2}$

This is intended as one of a few optimal models for solving the problem. The path analysis diagrams in Chapter 6 will illustrate that other paths, and therefore other rules, may be followed. However, the logic in another model would not be fundamentally different, even though there might be differences in direction at certain points. For example, R could equally well have been assigned before A. No new information became available through the assignment of A: it was simply that the information relating to R was not accessed at that point. This example illustrates that although a relationship exists between the letters in this module (Module 3: A,L,E/R,G), it is not one of extreme interdependence. Possible variations of this nature simply indicate the possibility for individual manoevring within certain boundaries.

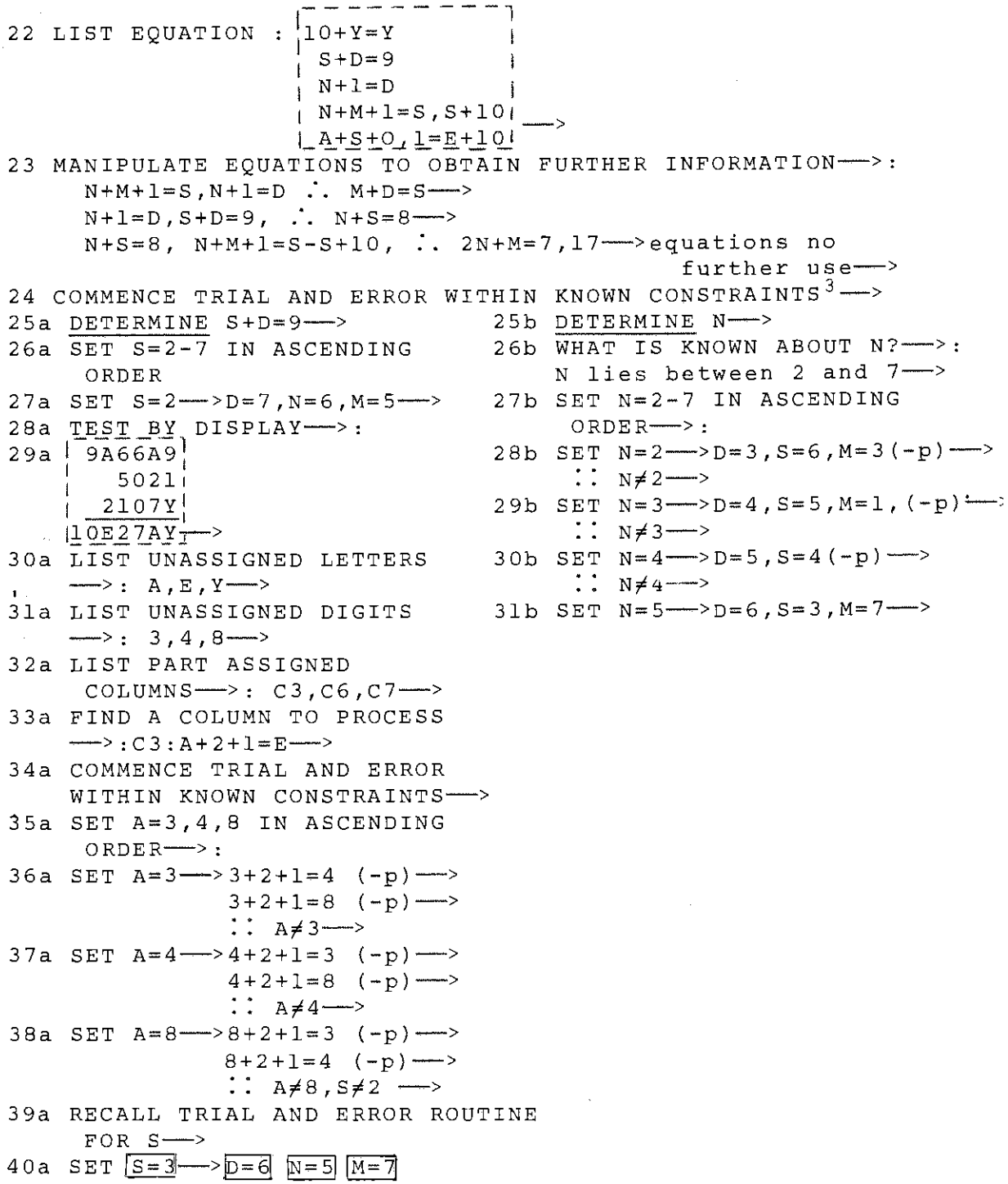
---

[2] Notice that the final display has a box with solid rather than dotted lines. This is because it represents the attainment of the final goal, and its goal properties transcend its function as a display of information.

(b) 'Ideal' Protocol for Problem 3

	C1	2	3	4	5	6	7
R1		H	A	N	N	A	H
R2				M	U	S	T
R3			S	T	U	D	Y
R4	T	U	E	S	D	A	Y

3.1 FIND AN ENTRY POINT⟶ :
2 WHERE IS MOST INFORMATION AVAILABLE?⟶ :C1R4 (T)⟶
3 PROCESS COLUMN 1⟶ : (T=0,1,2)⟶ (but T≠0, rule 3; T≠2, H+carry ≤ 19)⟶ : [T=1]
4 FIND A T⟶ :C7R2: H+T+Y=Y; 10+Y=Y⟶ : [H=9] ⟶
5 ANY OTHER T's?⟶ :C4R3: N+M+T=S⟶ : no information, too many unknowns⟶
6 ANY OTHER H's?⟶ :C2R1: H+carry=U⟶
7 WHAT DOES CARRY EQUAL?⟶
8 PROCESS COLUMN 3: A+S=E ≤ 16⟶ : carry 1 to C2⟶ : [U=0] ⟶
9 FIND A COLUMN TO PROCESS⟶
10 DO DISPLAY⟶ :
11 [9ANNA9 / MOS1 / S1ODY / 10ESDAY]⟶
12 FIND A COLUMN TO PROCESS⟶
13 PROCESS FROM RIGHT TO LEFT⟶ : PROCESS C7:10+Y=Y: Y=free⟶ : carry 1 to C6⟶
14 PROCESS NEXT COLUMN⟶ :C6: A+S+D+1=A⟶ :S+D=9⟶ : carry 1 to C5⟶
15 PROCESS NEXT COLUMN⟶ :C5:N+O+O=D⟶ :N+1=D⟶ : carry 0 to C4⟶ :(N≤7, D≤8, since D≠9)⟶
16 PROCESS NEXT COLUMN ⟶ :C4:N+M+1=S,S+10: carry 0,1 to C3⟶
17 PROCESS NEXT COLUMN⟶ :C3:A+S+O,1=E+10 :carry 1 to C2⟶
18 PROCESS NEXT COLUMN⟶ :C1,C2 already processed⟶
19 ASSESS DEGREE OF COLUMN ASSIGNMENTS⟶ : [fully assigned :C1,C2 / part assigned :C3-C7] ⟶
20 ANY FURTHER INFORMATION?⟶ :all unknowns lie between 2 and 8⟶
21 CAN CARRIES BE FURTHER SPECIFIED?⟶ : [carry 1 to C1,C2,C5,C6 / carry 0 to C4 / carry 0,1 to C3]



[3] Two main paths were followed from this point. They are both given since this will elucidate future discussion about module structure.

```
41 TEST BY DISPLAY[4]
42 ┌ 9A55A9┐
   │   7031│
   │  3106Y│
   └10E36AY┘ ──>
43 LIST UNASSIGNED LETTERS──>:A,E,Y──>
44 LIST UNASSIGNED DIGITS──>: 2,4,8──>
45 LIST PART ASSIGNED COLUMNS──>:C3,C6,C7──>
46 FIND A COLUMN TO PROCESS──>:C3:A+S+O,1=E+10
47 COMMENCE TRIAL AND ERROR WITHIN KNOWN CONSTRAINTS──>
48 SET A=2,4,8 IN ASCENDING ORDER──>
49 SET A=2──>2+3+1=4 (-p)──>
              2+3+1=8 (-p)──> :: A≠2──>
50 SET A=4──>4+3+1=2 (-p)──>
              4+3+1=8 (p) ──>
51 TEST BY DISPLAY──>
52 ┌ 945549┐
   │   7031│
   │  3106Y│
   └108364Y┘ ──>:: A≠4 (no carry to C2)──>
53 RECALL TRIAL AND ERROR ROUTINE FOR A──>
54 SET  A=8──>8+3+1=2(p)──>:: [A=8] [E=2]
55 TEST BY DISPLAY──>
56 ┌ 985589┐
   │   7031│
   │  3106Y│
   └102368Y┘ ──>
57 LIST UNASSIGNED LETTERS──>: Y
58 LIST UNASSIGNED DIGITS──>: 4
59 ASSIGN [Y=4] ──>
60 DO DISPLAY ──>
61 ┌ 985589┐
   │   7031│
   │  31064│
   └1023684┘
```

[4] The two paths reconverge at this point, and follow the same route to solution.

Before entering a discussion of the specific features of the protocols presented above, comments are made concerning some of the more general characteristics of cryptoarithmetic problems.

The habit of solving an addition problem from right to left is deeply ingrained. In a cryptoarithmetic problem, this habit is dysfunctional, and progress is not possible until the solver realizes that it cannot be approached as a conventional sum. Most subjects do not persist in exhibiting this Einstellung effect however, and rapidly realize that the repetition of letters and their special relationships offer clues to solution; solution of the problem thus requires the combination of conventional arithmetic rules with the special structure of the problem.

In his discussion of the properties of cyrptoarithmetic problems, Bartlett (1958,p.60) asserts that a solution means an acceptable translation from one language system to another. Difficulties arise from this transfer since the structure and sense of each system must be simultaneously maintained. This interpretation of the cryptoarithmetic problem is misleading. In a strict sense, the initial letter form of the problem does not constitute a language, since the letters bear no intrinsic relationship to each other, and could be replaced by any

other symbolic notation. The problem would be one of translation if the letters were related to the digits according to the structure of the alphabet. In this case it would be necessary to keep in mind the structural features of both languages, alphabetic and numeric.

Although it is often necessary to resort to methods that are not strictly deductive or inferential, the 'tree' structures presented in Chapter 6 (Figures 6.1 and 6.2) demonstrated that the 'guessing' that occurs at such points is never random. Some information is always available to the subject, so that when he makes the decision to replace his present method with trial and error, he never works randomly through all possible letter- digit combinations, but only through some subset, delimited by previously established facts and inferences. When the restrictions placed on the range of values possible for a letter are the outcome of a reasoned argument, as for example in the statement "R is odd, because 2L+1=R", then the trial and error is algorithmic, since it is guaranteed to produce the correct assignment: the trial and error routine would then test all odd numbers, and eventually confirm that R was equal to 7. On the other hand, if there is arbitrariness in delimiting the region, or if the basis of the decision

is a "hunch", as in statements like "What's a nice good looking number for D to equal?" or "Let's choose lower numbers, they're easier", then trial and error has heuristic properties.

We turn now to a more specific consideration of the salient features of the two protocols presented above.

(d) Problem 1:

```
  DONALD
  GERALD
  ROBERT
```

In Problem 1, the question of choosing an entry point does not arise, since one element (D=5) already exists in the problem space. The tendency in processing this problem is to move from column to column, until a column is found from which information may be extracted. The subject realized that E and R are the key steps, E more than R since E can be definitely assigned, independently of any other assignments being made, whereas R is logically uncertain until E has been eliminated. We have seen in the previous chapter that about twenty-five per cent of the subjects correctly assigned R before E, but it is important to note that this step was logically vulnerable. Once E is established, the other letters follow in a pattern of interdependence. Although in the ideal protocol the path goes from E to A, it could just as easily have gone from E to L, or from E to R. This is true of the other letters in the module (Module 3:G,L,A,E/R):

once a certain letter is established, say R, it is possible to go in several directions to process any of several letters or columns which are all in the same state of assignment (see Section 7.5).

After the first seven letters have been assigned, the subject resorts to trial and error. He has the digits 2,3 and 6 left to assign to the letters B,N and Ø. Here he is guided by the knowledge that there is a carry from Column 3 to Column 2, which means that N+7=B+10. It is a simple step to infer from this that N must be greater than 3, but most subjects simply manipulate the digits till some combination meets the requirements without contradiction. The mean time for this module was 3.06 minutes, and apparently presented little difficulty.

The need referred to earlier (Chapter 6.5) for distinguishing the between-module from the within-module phase will have become apparent: in the protocol for Problem 1 (DONALD) for example, the between-module processing between Modules 3 and 4 starts after the display has been evoked (Step 34) with the instruction "LIST UNASSIGNED LETTERS", and ends when trial and error reveals that '6+7=3 (p)" (Step 44). The within-module processing begins when 6 is assigned to N (Step 44) and ends when 2 is assigned to Ø (Step 47). The processing time between the last element of the module being assigned and the final

form of the solution being evoked as a display, is not part of the last module, but constitutes checking time for the entire problem. The problem (stage time) ends when the final display is confirmed as being the solution. The two phases of processing may be explicated in a similar manner for the other two modules of the problem.

The common properties shared by the two addition problems, as well as the unique features of each, will be clarified by the description that follows below of the solution patterns for Problem 3 (Since Problem 2 is a multiplication problem the 'ideal' protocol for the problem and a discussion of its features is postponed to p. 230).

(e) Problem 3: HANNAH
MUST
STUDY
TUESDAY

The preliminary activity in attempting to solve this problem consists in searching for an entry point. The structure is scanned, and it is very plausible that the pereceptual processes described by Simon and Barenfeld (1969) are in operation here. An analysis of the eye movements of the subject are likely to reveal a high speed scanning during the first few seconds of exposure to the problem, during which substantial information may be gathered. This is probably especially true of subjects with

previous experience of cryptoarithmetic problems. The outcome of this preliminary activity is the abandonment of the directional (right-to-left)strategy, and a location of the entry point to the problem as being situated in the extreme left-hand column. This is facilitated by the conspicuous appearance of the letter T, which is related to the letters H and U in such a way that once T has been decoded, H and U are the next to be solved (see protocol). However, while the search for the digit equivalents for T,H and U is in process, the subject is also processing other columns, and accumulating information about letters from other modules. This is partly because the subject is in fact not aware, especially at this point, that letters are grouped in certain ways, and partly because the column structure intermingles letters from diverse modules. In his search for information about T,H and U, the subject also finds out about other letters. This information is mainly couched in equation form. More accurately, much of the manipulation of relationships between letters results in pseudo-equations, since no true algebraic transformation is involved. This is more true for Problem 1 than for Problem 3. In the latter, some of the information is represented in genuine equation form (Step 23), and comprises an essential heuristic step in the solution process. This is more true of the letters in Module 2 than of any of the others.

The manipulation of equations is essential only as an intermediate step, and must be superceded by trial and error. Almost every subject states, in some manner, his realization that "logical" thinking is now unproductive, and that he must instead "try some things". As illustrated in the 'ideal' protocol above, this usually involves the testing of values for the letters S and D, according to the rule posed by the relationship S+D=9. Some subjects however, focus on N and derive values from the relationship S+N=8 and N+1=D. This will explain why Module 2 is in fact closer to a two-module situation. The reason that the $\hat{k}$ estimates for the module are less than 2 ($\hat{k}$ = 1.72) could be traced to the fact that there is no search time between the two 'sub-modules'. However, the four elements, S,D,N and M, can clearly be considered as two pairs, so that Module 2, in Figure 6.6, may be reconsidered in the following way: it becomes trivial to distinguish whether S or D was the first letter to be assigned, using the equation S+D=9. The distinction between those subjects who solved the equation by setting values for S, and those who set the value for D, is dispensable. The same reasoning applies to the second pair of letters, N and M, and to the two pairs of letters in the other combinations. When the paths of subjects are amalgamated in this manner, there are only 3 demonstrated paths for Module 2, instead of 12:

$\boxed{\begin{matrix}NS\\DM\end{matrix}}$ (N=8) $\boxed{\begin{matrix}DS\\NM\end{matrix}}$ (N=34) $\boxed{\begin{matrix}ND\\MS\end{matrix}}$ (N=7)

Although the four letters in the module are assigned in rapid succession, one quite distinct and independent concept is involved for each pair in the module. This provides support and explanation for the $\hat{k}$ estimate reported earlier (Table 4.10, p.126).

Once these four letters have been assigned, the same action reported for Problem 1 occurs: the subject sums up his current information, and asks what letters remain to be decoded. Both this situation and the response to it parallel that found in Module 4 of Problem 1. A short period of trial and error follows, and in each problem there is one 'free' letter (ø in Problem 1, Y in Problem 3) to which the last remaining digit is assigned. The stage times are remarkably similar: Module 4 (DONALD) = 3.06 minutes and Module 3 (HANNAH)=3.09 minutes. The processing time for this module is spent on (a) the accumulation of unused digits and letters and (b) the application of a trial and error subroutine to obtain satisfactory assignments. This differs from Module 2 where the subject does not specifically restrict his attention to the four letters of the module, and where the domain of admissible values is much larger. Consequently, the procedure is more lengthy and also more

prone to error. Most of the errors that occur in the problem arise during the search phase between Modules 1 and 2. The subject is either unable to recall previously assigned digits, is faulty in the application of methods, generates false assumptions about the carries into a given column, or simply produces computational errors. In the first problem, the errors are more likely to concern the carries than anything else. For example, by not realizing that L+L+1=R+10, the subject is not aware that L could be greater than 5. By not realizing there is no carry from A+A+1=E, the subject does not restrict A sufficiently (to be less than 5). Similarly with the key letter E. It cannot be solved until the subject realizes that there is a carry into the Ø+E=Ø column. Much time is spent on this column in an attempt to explain such a configuration of elements if E is not equal to zero.

Very few mistakes arose through either forgetting or misinterpreting the rules themselves. Only rarely, for instance, did a subject forget that each letter had a unique digit, although some subjects explicitly mentioned the rule that states that no left-hand letter can be equal to zero. This occurred mainly in the multiplication problem.

The effect of the first problem on the second (or third) was, apparently, remarkably slight. Mention was rarely made of methods used or situations encountered in the

earlier problem. It would be rash to conclude from this that previous exposure to these problems is unimportant - we have suggested in Chapter 4 that in fact such experience may be critical for differentiating subjects in terms of solution times. However, these speculations must take note of the findings reported by Hunter (1959) and Nance and Sinnot (1963) which were cited above (p.102).

(f) Problem 2:

$$\begin{array}{r} SIR \\ \hline I \\ \hline PEEP \end{array}$$

The general flow of the multiplication problem solution outlined in the previous chapter established the need for distinguishing between the three parts of the process, the details of which will now be elaborated. Notice that the paths to solution may be readily and parsimoniously expressed in conventional flow chart form.

Problem 2

	C1	2	3	4
R1		S	I	R
R2				I
R3	P	E	E	P

2.1 HOW MANY LETTERS?——>:
2 5: (I,R,P,S,E)——>
3 WHAT INFORMATION IS AVAILABLE ABOUT LETTERS?——>:
4 I,R,P,S,≠0; I,R,P≠1; I,R,P≠5——>
5 ANY FURTHER INFORMATION?——>:
6 If I even, P even; P is less than I; I=2,3,4,6,7,8,9; carries=0-7 ——>
7 FIND A COLUMN TO PROCESS——>: C1:RxI=P
8 No meaningful restrictions, domain too broad, no processing possible——>
9 FIND ANOTHER METHOD TO REDUCE POSSIBLE VALUES——>:
10 SET UP EQUATIONS——>:
11 SIR=I(R+10I+100S)=P(1001)+E(110)——>

```
12 PRODUCE SMALLER EQUATIONS——>:
13 IR=P+10x; II+x=E+10y; IS+y=10P+E——>
14 FIND A COLUMN TO PROCESS——>:C1:RxI=P——>:
15 SOLVE EQUATION IR=P+10x——>:
16 Too many values, not enough constraints, no processing
    possible——>
17 FIND ANOTHER METHOD TO REDUCE POSSIBLE VALUES——>:
18 COMMENCE TRIAL AND ERROR WITHIN KNOWN CONSTRAINTS——>
19 FIND A COLUMN TO PROCESS——>:C1:RxI=P——>:
20 FIND A LETTER TO PROCESS——>:I:C4R2——>
21 ESTABLISH SUBROUTINE——>: SET I, SET P, FIND R, FIND E,
    FIND S——>
22 SET I=2,3,4,6,7,8,9 IN ASCENDING ORDER——>:
23 SET I=2——>(-p: P evenP<I) ∴ I≠2——>:
24 SET I=3——>P=1,R=7,E=1 (-p:E=P)——>
              P=2,R=4,E=0, S (-p:Sx3≠19) ∴ I≠3——>
25 SET [I=4] ——>Peven,<4: P=2,R=3,E=7  S(-p:Sx4≠26)——>
                          [P=2,][R=8,][E=9,][S=7]  (p)——>
26 DO DISPLAY——>
27  748
      4
   2992
```

In the beginning, the subject scans the problem as a whole, searching for ways to define it. A typical comment at this stage is:

"I wonder if there is some significance to PEEP". He also searches for methods that would be of use in solving the problem. The methods generated by subjects vary, and most protocols include more than one. The principle heuristic methods are used to limit the admissible range of values for a letter, or for all the letters. This may be expressed as inequalities, "P is less than I", or vectors "I must be even, 2,4,6,8" and "I must be a prime", or by defining the range of possible values for a carry.

Another method that appears is the construction of equations. In this problem, unlike the two addition tasks, the equations are genuine algebraic statements, but this method is soon abandoned when it is realized that there are too many unknowns to allow the equation to be solved. When all these attempts fail, the subject realizes that the problem can only be solved by the application of trial and error, or in the words of one subject by "systemmatic variation to produce maximum restrictions". This decision is usually arrived at reluctantly, since the prospect of routine setting and resetting of values is a tedious one. At this point, there is frequent reference to the fact that previous methods failed:

> "Factor 11 for R implies R+S=1 or -11. Factorizing turned out to be a blind alley."

Many subjects appear to think that the use of trial and error is an admission of defeat, that it is not a legitimate method but represents instead an inability to find a more logical line of attack.

Once the decision to use trial and error is made, the subjects search for a letter on which to focus. This is usually I, the multiplier, although sometimes P is used, and, very occasionally, the search is focussed on a list of perfect squares to settle I^2. This last attempt

has to be supplemented with additional information, since, without knowledge of the carries involved, the information is ambiguous. Almost every subject sets the values of I in ascending order, saying for example,

> "Calculations are easier from the lower end, so unless there are reasons to start higher up, it is better to start lower down."

For this reason, of four possible solutions, the one most frequently attained is the one illustrated in the protocol above. Alternative solutions entail a higher value for I. The majority of subjects are methodical in their application of the subroutine for setting the values of the letter concerned, but an occasional individual will select values haphazardly, saying:

"Let's try 7, I like 7."

The restrictions imposed on the set of possible values are often forgotten - it appears to be cheaper, when calculations are relatively simple, to test all values, rather than to retrieve and store in immediate memory a certain rule, such as 'P is even'. This processing subroutine is set out in Step 21 of the 'ideal' solution to Problem 2 (p.231). Briefly, in this second 'chunk' of processing, the subject fixes on values of I or P by varying this relationship until a solution is found.

The third and final part of the problem solution consists of entering the module itself, that is, the assignment

of the correct digits to the letters in the problem. As with the other problems, the values are checked by calling the display in order to check that the assignments that have been made are compatible with the solution.

There are some interesting features that distinguish the multiplication problem from the addition problems. For example, since there is more dependence between the letters of the column (and problem) than in the addition problems, greater emphasis is placed on the column as a unit of processing. A letter is assessed entirely in the context of its relation to the other column elements, since it is not possible to delimit the domain of values possible for a letter by any other means. Another obstacle to placing bounds on the processing region is the large range possible for the values of the carries in and out of the columns compared to Problems 1 and 3 where the range is theoretically restricted to the values 0,1 and 2. A further innovation is the attention paid to the units digits of the product row. The use of the planning heuristic is evident in the protocols: planning involves creating a simpler model of the problem which is then treated as the object to be solved. Examples of this are the attempt to classify the problem as a palindrome, or as a division problem, or to classify the multiplier as a prime. There is reason to assume that the subject has available a multiplication

tableau'in his head' to which he makes frequent reference by either list searching or direct assessing methods (see Section 7.6I). There are many questions of the form 'What times what equals x?' and the subject may be viewed as searching through the multiplication tableau till he finds a combination of digits that satisfy the question.

Although the stages-module model did not provide an effective description of the multiplication solution process, there are abundant features of interest to be found in the protocols which encourage the future use of other multiplication and division problems.

7.2 Verbalization

This study was not designed to examine the effect of verbalization per se, and it will therefore not be possible for this section to present more than an impressionistic analysis of the effects and mechanics of verbalization. It is hoped that future investigations will be directed to some of the questions raised in Chapter 1. The development of a body of research around the process and effects of verbalization is not only valuable in its own right, but essential within the context of the New Look in computer-based psychology.

The first impression to be gained from reading through the protocols is one of variability, which may be measured by two indices: (1) rate of verbal output and

(2) regularity of verbal output. Protocols were randomly chosen for closer, but necessarily cursory, examination. The one-minute intervals used for investigating stage times provided the temporal unit of measurement, while the verbal variable was defined in word units. This is possibly a more meaningful measure than the syllable, since, in long words especially, there is a strong tendency for syllables to be shortened.[5] The rate of verbal output is indicated by number of words per minute, an average rate calculated by the number of words spoken over the whole protocol.[6] Regularity of rate is indicated by the highest and lowest rates occuring in any one-minute interval.

Large individual differences appeared in both these variables. For example, for one subject, the maximum and minimum rate was 70 and 50 words per minute respectively, a high and regular rate of performance. Another subject, displayed an equally regular but slower rate, with a maximum of 42 and a minimum of 35 words per minute. On the other hand, there was a high degree of irregularity, as well as extreme rate values in the performance of the

[5] Although this may be disputed by linguists. J.Bernard (personal communication) prefers syllable counts since a subject who used many polysyllabic words might score a lower rate assessment than one who used few, even if the actual output of the latter were lower.

[6] It is not possible to compare this to standard Australian utterance, since available results are expressed in syllables per minute (Bernard, 1965).

subject whose high rate was 105 and whose low rate was 19 words per minute. It seems unlikely that either relative rate or regularity is related in a simple fashion to facility in problem solving. Of the three subjects described above, the first solved Problem 3 in 11.62 minutes, the second solved Problem 1 in 43.58 minutes, and the third solved Problem 3 in 15.40 minutes. (However, it is possible that it is not the actual difficulty of the problem which is critical, but the length of time over which the verbalization process is active. This question can only be investigated through an appropriately designed experiment.)

An analysis of overall rate substantiated the impression that no obvious relationship exists between verbalizing and experienced difficulty in solving the problem.[7] Two subjects, who each solved Problem 3 in about 11 minutes, performed at an absolute verbalization rate of 60 words per minute, while a subject whose solution time was 22.50 minutes, verbalized at about 20 words per minute. Other subjects, also with solution time of about 22.50 minutes, varied from 46 to 75 words per minute. Since this is an overall rate, these figures reveal substantial individual differences in verbal output.

---

[7] Difficulty is assumed to be linearly related to solution time (see Chapter 4).

However, a complete description should include the fluctuations displayed within an individual protocol.

We have seen that a subject may generate a regular or irregular flow of verbal output, and he may be voluble or reserved irrespective of whether he finds the problem easy or difficult to solve. This is substantiated by the differential response to local difficulty: one subject will lapse into silence while another may increase his rate of talking. The same situation exists when a subject is about to achieve success, or when he is involved in computation. For some subjects,it is conceivable that expectation of success acts as a signal for lowering the threshold for verbalization, while others apparently fear that talking will disturb a hypothesis that is incompletely conceived or tested (Phelan, 1965). It is plausible that some threshold must be traversed, and that there is considerable individual variability in this. A concept like Polya's (1954) "index of plausibility" is useful for expressing the rationale for an individual decision to utter a verbal statement. This may or may not be subject to vacillation. In the subjects whose maximum and minimum rates do not differ to any great extent, the index of plausibility remains constant. For other subjects, increased confidence or motivation could act to alter the decision rule that determines the level of the threshold for verbalization.

Can a 'personal' rate be detected? (This should not be confused with λ although it is conceivable that some connection could be forged between the two.) There is evidence that, for longer utterances of the order of 100 syllables or more, the rate of utterances tends to be constant enough for an individual that it may be considered a personality constant (Goldman-Eisler, 1954).

The evidence from the protocols is not so unequivocal, but it must be remembered that neither the tools of measurement nor the experimental techniques were designed to investigate this phenomenon. For one subject who solved all three problems, there is striking similarity between his performance on two problems: on Problem 3, which he required 15.40 minutes to solve, his maximum and minimum rates were 105 and 19 words per minute respectively. On Problem 2, which he solved in 17.00 minutes, his rates were 110 and 19 words per minute. However, on Problem 1, which he apparently found very easy, solving it in 6.58 minutes, his maximum was 107, and his minimum 57 words per minute. The maximum rate remained stable, but the minimum rose, a reflection of a more homogeneous and effortless performance. There is further evidence from other subjects: on two problems, one subject displayed rates of 57/100 and 66/125 words per minute, another 7/70 and 15/70 and another 39/100 and 43/99. The ratio of minimum to maximum rates varies widely

between these individuals, but there is evidence for a consistent verbalization style between problems for the same person.

There are large qualitative differences in the protocols. Although some subjects are extremely voluble, the content of their protocols consists largely of problem-relevant utterances. Other subjects may be much more informal, and introduce many irrelevant remarks, such as commenting on the earphones being worn, the time, the person behind, etc. A few protocols consist of very sparse output, mainly in the form of digits and letters, with little reference to methods. For example, a typical section of such a protocol reads

> "N is 6,T is 1, E,what'sE, anything, we have 3, make it 4,make it 5, 12, 6,8, makes it 15, make it 1, makes it 2, makes it 9, 5's are 12, 6, no good..."

and would yield little of interest when analysed. This subject had a generally low and irregular output rate, with a minimum of 3 and a maximum of 49 words per minute, and an overall rate of 30 words per minute. Incidentally, this rate of one word every two seconds is similar to the subject (S3) reported by Newell (1967).

Although the available numbers are very small, it is worth noting that five of the six female subjects in the experiment had extremely high and regular rates of verbalization. The exception was a computer programmer who produced

the two shortest protocols in the study (see Appendix I2 and I9).

There was very little reported difficulty in verbalizing - indeed only two subjects specifically referred to the fact:

> "You people out there who are listening to this, I hope you know I'll be feeling like a nut. I feel like a nut talking to myself."

> "I find this talking actually slows me down a bit, but never mind."

For the second subject, his times to solution for the second and third problems were well below the mean, whereas for the first it was above. It is likely that verbalization did in fact slow him down in the beginning, but that, with subsequent practice, this effect was neutralized.

Most subjects appeared relaxed, and the fact that several informal activities occurred, such as whistling, speaking French etc., provides encouraging evidence that the process of verbalization is not a hindrance but a close reflection of thought itself.

It is clear from the foregoing discussion that the process of verbalization is a topic rich in research hypotheses. We assume we are tapping an internal space, but the rules of transformation from internal to external space cannot be specified in the absence of controlled

studies. Ideally, such research should proceed in synchrony with analyses of nonlinguistic material such as eyemovements, and a linguistic analysis of pauses, volume, pitch, etc. Protocols are, it is claimed, reliable for the information they contain, but considerable essential material is omitted by exclusive reliance on verbal behaviour. Further questions that warrant answering are whether a relationship (possibly curvilinear) exists between verbal rate and beginning or end of problem. It seems likely that the rate would be higher at these points, where there is less density of processing activity. At the beginning of a problem there is scanning, rephrasing of the problem, ennumeration of its elements, and at the end a phenomenon described by J.R.Hayes'(1966) 'solution acceleration' or Hull's (1943) 'goal-gradient' seems to be in operation. Is there a curvilinear relationship between difficulty and amount of verbal output? We have suggested that for some subjects verbalization increases while there is a lull in information processing. Perhaps for these individuals verbalization is a constraining activity, so that there is an inverse relationship between thinking and verbalizing. We might also inquire into whether a slow rate of verbalizing indicates that thinking is also proceeding at a slow pace, or alternatively, whether verbalizing is a nonisomorphic model of thinking in which speech trails

thought, sampling and summarizing it at certain junctions.

7.3 Newell's Meta-language for Protocol Analysis

Equally important, but proceeding in a research space distinct from investigations into the process of verbalization itself, is the search for tools with which to analyse the products of that verbalization.

The two papers (Newell,1966a, 1967) discussed in this section are an important contribution to the search for a methodology of protocol analysis. It is not possible to apply these methods wholesale to protocols gathered in this study, since the production system language was developed by Newell with a different objective in mind, namely that of building a computer program to simulate the solution of cryptoarithmetic problems. The concern of this thesis is with description at a grosser level, and with the understanding of major patterns of behaviour. This difference in level of investigation is reflected in the fact that the Newell study focusses on the behaviour of a single individual, whereas our interest concerns the similarities and differences that may be found over a range of individuals. Despite the divergence of emphasis, the system developed by Newell will be described at length in this section, since it supplements and enlarges our perspective of the problem - solving process.

Newell's meta-language may be useful for the analysis of chess problems as well as for cryptoarithmetic problems despite large differences in the problem spaces (see Chapter 1). For example, in chess, the subject revealed an overall search strategy (progressive deepening), whereas no such consistency appeared in the behaviour of the subject solving the cryptoarithmetic problem. Newell points to the possibility of a theoretical framework large enough to incorporate both types of process. He also suggests that GPS could be reformulated to merge with this system, so that the production system forms a quite general form of process organization.

Newell's results may be accused of being idiosyncratic, since the entire meta-language apparently derives from the analysis of a single protocol. He asserts, however, that the rules were in fact developed in conjunction with "various notions" of how cryptoarithmetic problems are solved (1967,p.56). There is no reason to dismiss as 'unscientific' the examination of one's own experience as a fruitful source of hypotheses, as long as these hypotheses are then subjected to experimental test. General confirmation of the mechanisms postulated by Newell comes not only from this study, but from Bartlett (1958) who gave Problem 1 (DONALD) to his subjects, instructing them to write down their steps to solution. Again, the levels of

analysis intended by Bartlett and Newell differ widely, but the general flow of reasoning generated by their subjects and those of the present study show substantial agreement.

The rationale for a detailed account of Newell's methodology is provided by the following quotation:

> Another issue concerns the psychological significance and generality of our productions. If this model represent structurally what is going on, then the individual productions must be learned, transferred etc. Are the same productions used in other tasks? Do the same productions show up in different people? Here it seems to me, we are in batter shape than we have any right to expect. A number of productions are clearly of general utility....most of the others could be so reframed. (1967, p.130)

Our intent here is to provide support for the above claims. The operators and productions posited by Newell appear to be sufficiently fundamental to describe the behaviour of diverse subjects on another addition as well as on a multiplication problem. It is true that a few subjects attempt the problems in totally different ways; some, for example, rephrase the multiplication problem as a palindrome, or formulate a global rule such that digits are assigned to letters according to alphabetic rules. However, these do not usually lead to successful solutions, and subjects who use them either fail to solve the problem, or abandon the method in favour of a more conventional one.

The process of problem solving is analysed by Newell in the context of operators and productions. The four operators are defined in terms of input and output variables; the productions are divided into four classes on the basis of their functions. The definitions are taken directly from Newell (1966a, p.12a; p.27-29):

Operators

(1) PC(c) Process the column c. The input is all the information about the column, including the letters and carries in it; the output is some information that can be inferred from the column.

(2) GN(v) Generate the values of variable v.

(3) AV(v) Assign a value to the variable v. The output is in the form of a digit assigned to a letter. The value is selected from the set generated by GN(v)

(4) TD(l,d) Test if a letter can take the value of a certain digit.

Productions The production system is a language to express the decision and selection processes used by the subject. This has been described in Chapter 1, but a brief restatement is in order here. A production system consists of a set of productions, each of which consists of a condition expression followed by an action expression:

condition ——>action

Since the production is to be considered in the context of the state of knowledge at a node, then

cues in knowledge state——>action

is a conditional expression, since the occurrence of the action (one of the operators along with its operands) depends on the occurrence of the cues. This will become clearer with the aid of illustrative sections of protocol. The four types of productions are:

(1) S1-S5: These lead to the selection of one of the operators PC,GN,AV.

(2) G1-G5: These provide mechanisms for setting up a goal, either to get something, or to check something.

(3) T1-T2: These productions are concerned with terminating lines of search. They govern the selection of operator TD.

(4) R1-R2: These are concerned with repeating previously trodden paths.

The remainder of this section will attempt to clarify the way in which the system is applied to a protocol, by using examples from the protocol of the subject (S3) described by Newell. The segments will be selected for comparability with similar situations found in the protocols of subjects from this study working on other problems. By these examples, it is intended to show both the level on which Newell is working, and why the techniques proposed by him are, at the same time, too detailed and too

inferential for the purposes envisaged by the present work. Our concern is not to discover or infer what is occurring during a particular period of silence, nor to provide bridging mechanisms for places in the protocol where the subject has not made explicit his line of reasoning. For Newell, such a rigorous and detailed procedure is necessary if he is to construct a working program that will solve cryptoarithmetic(and other) problems. To develop his theory, it would be necessary to embark on an equally intensive analysis of one or more subjects on the same or other problems, and such a scheme is clearly far beyond the scope and intent of the present study. It will be sufficient for us to demonstrate that, although any investigation of cognitive processes using the protocol technique would be incomplete without a detailed mention of Newell's work, it also represents a tool too fine for the objectives of this research.

(The notation used is that of Backus Normal Form (Backus, 1959), which is convenient for constructing schemes for the class of expressions representing states of knowledge, and then for defining the set of operators in terms of these expressions. Those symbols that appear in the examples below will be defined in context.)

Examples of productions and operators

(1)	B78 Now I have this Ø repeating here in the second column from the left; B79 that is, itself plus another number equal to itself. B80 This might indicate that E was zero.	S2: get t1→ FC(t1) =>c2;PC(c2,t1) =>E=0
	B80.1	R1: PC unclear→ get E;repeat PC
	B81 In fact it might have necessarily to indicate that	↑:PC(c2,E)=>E=0
	B82 I'm not sure	R1:PC unclear → get E; repeat PC

The phrases labelled B... "are based on a naiive assessment of what constitutes a single task, assertion or reference." (1967,p.31) Breaking the protocol into these elements provides some form of measurement of what information the subject had at a particular time. The lines drawn across the protocol constitute a node or a new state of knowledge. This may consist of one or more phrases. In B80.1, the empty section refers to the fact that some action occurred between B80 and B81 which needed to be specified by the analyst, since there is no verbal evidence of the action. Newell would have to defend the use of inference of such situations against the charge of

extravagant interpretation of the contents of the 'black box' he is describing. The statements B78-80 are referred to by:

S2: get→t1 FC(t1)=>c2;PC(c2,t1)=>E=0

The condition part of a production occurs on the left of the arrow (→) and the action part on the right. The action may consist of a sequence of actions, separated by ;. The double arrow (=>) indicates the output of a process. The selection production (S2) is evoked here: the goal is to get information about some variable, in this case the carry into column 1 (get t1). This then leads to the process (not an operator, since it does not lead to a change in the state of knowledge) of finding a column (FC) which contains that variable (t1), which produces column 2 (=>c2). Then process that column for the variable (PC(c2;t1)). PC is one of the operators, and its output is the statement "this might indicate that E is zero" and is expressed (=> E=0).

The production S2 then, describes the condition, as expressed by the cues existing in the state of knowledge: the goal of getting information about the carry into column 1; this is followed by the action part of the production - finding and processing the appropriate column.

In B80.1, it is inferred that the repeating production (R1) occurs, which is evoked when the result of a process is unclear (PC unclear). The action part of the production consists in obtaining information about the variable E (get E) involved in the unclear statement. The production then involves repeating the process which was unclear.

In B81, the process is repeated (the upward pointing arrow refers to the fact that the same production is evoked as at the preceding node, B80.1) and the operator PC processes column 2 to obtain more information about E (PC(c2,E)), the output from which is again E=0. In B82, the production (R1) evokes the repetition of previous paths.

We will now provide examples of similar situations and of other subjects that can be handled with the same language. For the sake of simplicity, the segments of protocols will not be broken up into smaller units, and the series of productions will describe the entire example:

"Now can we fix T perhaps? Under this assumption, I haven't carried 1 from N+M+S, N+M+T rather, so A+S could be at worst, so the worst I could carry from A+S is 1."	S2: get T→FC(T)=>c4;PC(c4)=> N+M+T=S ↑: get c5;PC(c5)=>t5=0 ↑: get c2→FC(t2)=>PC(c2)=> t2=1

A second example,without being expressed in terms of productions, describes a processing situation identical to the one occurring in the illustration from Newell's subject:

> "We see here that Ø+E gives Ø,er, now this means that Ø ...yeah, I think that means that E must be zero. Only number added to itself gives another number is zero. If Ø+E gives Ø, or Ø+E+1 gives Ø."

It is clear that similar production expressions will describe this segment.

Although the language we have been detailing is an excellent one for microanalysis, a later section of this chapter will develop a more general scheme for describing the behaviour, both integral and auxiliary to the solution process itself, exhibited by subjects through protocols. The micro-approach results in a highly localized view of the problem solver . When the aims are more general, and the intention is to incorporate as much of the data as possible, a more appropriate procedure is to step further back so that grosser aspects of the behaviour can be observed.

7.4 The Stages-Module Model and Information-Processing Concepts

It has been asserted (Newell, 1966b, pp.175-176) that any stage-based model is unable to capture the nonlinear, search quality of problem solving. Most conventional stages models deal with aspects of the problem solving process such as incubation, inspiration, preparation etc., and it is perhaps unforunate that the same terminology should be used to refer to the model on which this thesis is constructed.

We have seen that the module concept provides a useful model for the sequential assignments of digits to letters in cryptoarithmetic problems, and that the times to solve a module are efficiently handled by the stages model. Can one go deeper, and ascertain whether the stages-module model remains a viable one for describing recurrent mechanisms to be found in the protocols? It may be, for example, that characteristic operations occur at the beginning of every module. Are there necessary features that appear in every segment of processing (including between-module phases) that culminate in the completion of a module? In other words, does the module model provide a template that can be placed unambiguously over those parts of the protocol which we have been calling modules, that is, the part that begins at the

end of one module and continues to the end of the next module?

The answer to these questions is disappointing. Only very rarely does a subject display any awareness of the modular structure of the problem. It is rare to find remarks such as:

> "All the first part of the problem would be filled in...and then the next part of the problem would be ..."

and

> "Put these in a separate group, D,N,S,M."

The typical subject does not indicate that he knows he must solve, or has solved, the letters in groups. However, there is evidence that lends suggestive, though not very substantial support to the possibility that a module has significance in information-processing terms. This argument will now be developed.

There is a widespread tendency among the subjects to call a visual display (see Section 7.1) manifested as a tableau of accumulated information. Ideally, this is reconstructed in the form in which the original problem was presented, but with known assignments substituted for letters in the columns. Often, however, the reconstruction is only partial, so that only a part of the tableau actually appears. Although every protocol contains examples of this display evocation, it is less common to find individuals

who use displays only after the completion of a module. It is more usual for a subject to use a display after only one or two modules, although occasionally it appears after the assignment of every letter. The most frequent occurrence of the display mechanism is to be found at the end of Module 2 in Problem 3 and Module 3 in Problem 1 - in other words before the beginning of the last module, when all current information is retrieved in order to assess the distance remaining to the goal (mean-end analysis):

> "S is 3, and the carry is 4. A+4=3. Now what do we have left at this stage (sic!)?0,1,3,4,5, 6,7,8,9, U is 0, T is 1, haven't assigned 2..."

Here the subject is clearly filling in a tableau to establish what information is certain, and therefore what still remains to be completed.

Sometimes displays are evoked when the solver finds himself in difficulty. This may serve as a way of avoiding periods of silence during verbalization, or it may provide positive reinforcement by visibly demonstrating that some progress has been made either since the beginning of the problem, or since some particular point.

Often, the subject only produces part of the tableau, but this is usually associated with checking behaviour, when only the relevant part of the problem is processed. This can be seen in the following example:

"L=8. Now how does that change it? DONALD.
We've got the 5,2,D,G,L=8,5,5's 10,8,8,16."

It is clear that no watertight relationship can be drawn between the use of displays and the module structure of the problem, since there is wide variation in the completeness of the tableaux when they do occur, as well as individual differences in the points at which they occur - after each letter-digit substitution, after every module, and only then, after some modules and not others. It would nevertheless be misleading to assume that problem solving can operate in the absence of some kind of tableau-displaying device. It is essential that some form of evaluative mechanism exist, so that a check can be maintained on previously translated letters, and an assessment be made of the distance remaining between present and desired situations. A tableau display would appear the most efficient method, but the protocols do not yield as frequent examples as one would expect. It seems extremely likely that some form of representation must occur, either internally (and therefore, more subject to error), or externally, in either spoken or written form. The last possibility seems the most plausible, and is substantiated by the working paper used by the subjects, which, when scrutinized, disclosed varying amounts of information not contained in the protocols. This is a clear illustration of the shortcomings

of protocols: in this situation the protocols did not constitute a complete record of the ongoing process, and critical material was contained in some other problem space.

Although the data on tableau displays is equivocal and although for many subjects the behaviour that emerges after the assignment of any given letter is no different from that occurring after the final letter of a module (in both cases, the subject moves on immediately to find a new column to process), the tentative assertion is made that the module model does have some corroboration at the information-processing level.

The difficulty in conceiving of the model as a template to be placed over sections of the protocol with the aim of discovering recurrences of behaviour within it, is due to the fact that problem-solving behaviour is not a linear sequence of actions. While on the way to a certain module, the subject is spending time on letters from other modules. His processing paths may terminate for any of a variety of reasons, and may or may not be returned to later. Although a module is conceptually rigorous because its objectives are modest - it only accounts for the correct assignments of digits to letters - it has, by the same token, limited explanatory power. Modules, as we have defined them, are both too superficial and sequential to account for the nonlinear search quality

of problem solving. Although it is possible to theorize about the recurrent actions that occur - there must be recurrences at some level, since the procedure of assigning a digit to a letter is repeated, for example, ten times- this is not the most fruitful position to take. It is more valuable to treat the protocol in its entirety as a continuous stream of behaviour, and to observe actions without relating them to temporal or structural factors. This is not to deny that the description of some behavioural elements is enhanced by associating them in a time context, as for example the elimination process that occurs most commonly towards the end of a problem, or the displays that occur after a letter or module has been assigned.

A naiive interpretation of the stages-module model will clearly not contribute greatly to an understanding of the problem-solving process. On the other hand, it is not possible to dismiss the convincing stage ($\hat{k}$) and $\hat{\lambda}$ results presented in Chapter 4. It cannot be denied that the $\hat{\lambda}$ values make intuitive sense, especially the fact that these vary for different modules, the easier ones having higher probability rates. We may not discover an isomorphic relationship between structural aspects of the problem and the information-processing functions of the protocol, but we can look closer at the problem solver, and attempt explanations for his behaviour, paying special

attention to his response to the modular structure of the problem as this is manifested in the paths followed and the times taken.

Hypothesis and information pools

Problem 3 proved more difficult than the other problems not only as measured by mean time to solution, but also by a lower probability rate: $\hat{\lambda}$=.07 (Problem 3), compared to $\hat{\lambda}$=.09 (Problem 1) and $\hat{\lambda}$=.14 (Problem 2).

Two sources of this greater difficulty stem from the fact that in this problem ten letters had to be decoded, as opposed to nine and five in the other two problems. More important, however, are the larger number of situations to be encountered, a property of the problem discussed along the dimensions of elements and configurations in Chapter 6. What is the behavioural response to this greater complexity? The discussion that follows will develop a simplistic model of the situation, along with a keen awareness of its conjectural and untested nature.

Let us view the subject as testing various hypotheses and strategies about the solution to a problem. Each problem is defined by its own associated population of hypotheses. The sampling of hypotheses may occur with or without replacement. In the case with replacement, hypotheses that prove useless are excluded from the population as they are sampled, tested and rejected. If no new hypotheses are

added, the proportion of correct strategies increases with each error. In the case of sampling with replacement, the subject samples a hypotheses, and, whether or not it is applicable, he replaces it, and may or may not resample it. This implies that he has no memory of which previous hypotheses have been tried, and whether they proved successful or were rejected. The larger the number of hypotheses in the hypothesis pool, the smaller is the difference between sampling with and without replacement (Bower and Trabasso, 1967). In other words, where the pool is large, the constant probability of a correct hypothesis being drawn in the sampling with replacement situation is not substantially different from the increasing probability of sampling without replacement. We will assume that the pool contains a sufficiently large number of hypotheses to ignore the question of whether the situation is one of sampling with or without replacement. It seems likely however that the sampling with replacement case is a more veridical representation of the situation, since there is a high degree of forgetting of both previously successful and rejected methods, hypotheses and assignments. In this model, both successful and unsuccessful hypotheses are returned to the pool, but with a 'tag' labelling their success or failure in a previous situation. Errors in problem solving arise when the subject

loses or forgets the 'tag' he has used. This can result in a successful hypothesis being forgotten, or not selected in the appropriate situation, or the failure of a hypothesis is forgotten, and so it is reapplied. During sampling, any hypothesis may be drawn. (It is obvious that any development of this model would have to include specifications concerning the different probabilities that might characterize hypotheses, so that, for example, there may be more samples of a well-tried hypothesis in the hypothesis pool, which would increase the probability of that hypothesis being sampled, or selection might be influenced by a simplicity bias (Gyr,1960). Alternatively, it may be sufficient to assume constant probability for all hypotheses. Such considerations do not warrant careful discussion at this time.) Subsequent activity depends on whether it was applied or returned to the pool.

Independently of the above remarks, we continue to represent the subject as sampling from a <u>hypothesis pool</u>. Assume that the more complex task has associated with it a larger pool of hypotheses. Assume also a second concept, the <u>information pool</u>, which is empty at the beginning of the problem. The information pool is filled solely through the process of problem solving, and does not include knowledge that the subject might bring to the

problem. We have not made any assumptions about the probabilities attached to the sampling of a correct or an incorrect hypothesis, but it may be assumed that the more difficult a problem is, the more processing situations or elements it contains, and therefore, the larger is the attendant hypothesis pool. This leads to a smaller probability that an appropriate hypothesis will be sampled to match a particular situation. The subject therefore experiences increased difficulty because (1) there are more hypotheses to sample from, and (2) there is a lower probability of matching hypothesis with situation.

There is an interaction between the hypothesis and information pools, but most of the traffic goes from the hypothesis to the information pool. Just as each problem has a population of hypotheses, so does each module. These do not, however, exist but must be generated by the solver. The hypotheses are generated in accordance with the demands of the situation, and since the circumstances and requirements vary from module to module, a new hypothesis pool has to be generated at each new module. Specifically, the subject is viewed at the end of each module as ready to embark on the processing of the next module, and as generating hypotheses according to the demands of the situation, as he sees it. It is obvious that individual differences will arise in the acuity with which the

requirements of the case are preceived, and consequently in the sufficiency of the pool of hypotheses that are generated. There are (unspecified) learning and transfer effects from earlier hypothesis pools.

By what mechanism are the requirements of the new situation evaluated, and in what way is the cognitive system alerted to begin generating a new hypothesis pool? Both these actions are hypothesized to occur through the display mechanism. When a module has been processed, the subject stops, reviews his previous output and his state of progress by means of a display, and then prepares for the next phase by generating a relevant pool of hypotheses. For example, prior to entering the second module in the HANNAH problem, he surveys the number of letters still to be assigned. He also has earlier information accumulated during the pre-Module 1 search phase, and, confronted by the low degree of specificity of information (in the information pool) plus the large number of letters (7) still to be assigned, he generates a large hypothesis pool. By the time he enters the search for Module 3, he inspects the amount of definite information, and, by evoking a display, estimates that only a fraction of the problem remains to be solved. Consequently, he generates a much smaller hypothesis pool.

This interpretation is related to λ in the following way. If we simply posit one undifferentiated hypothesis pool for the entire problem, and if the situation is simply one of constant sampling with replacement, then λ would be expected to remain constant over all stages of the problem. Instead however, we observe different rates for different stages, and these correspond to the difficulty, as indexed by time to solution, experienced by the subjects; for example, the 'easiest' stage in Problem 3 was solved after a mean time of 3.10 minutes with $\hat{\lambda}=0.44$, while the 'hardest' stage required a mean time of 22.62 minutes with $\hat{\lambda}=0.08$, a lower probability rate. ($\hat{k}$ and $\hat{\lambda}$ values are presented in Table 4.12.)

At the beginning of the problem, the information pool is empty, and (for Problem 3) the hypothesis pool is not very large. Since no known information has accumulated through processing, there are no constraints imposed by the information pool. Considerable information is conveyed by the initial scanning of the structure of the problem, and, more specifically, by the conspicuous positions of the letters H,T and U, so that it is not necessary to generate a large hypothesis pool at the beginning of this problem. Those individuals who are not aware of the information conveyed by the structure of the problem will of course generate, and consequently have to sample, a far

larger hypothesis pool, which will lead to an increased solution time, and an increased value of $\hat{\lambda}$. At the end of the first module, the time required to search through the information pool is relatively short since the amount of information it contains is not excessive. As processing continues between Module 1 and 2 however, information accumulates rapidly, in the form of equations and vectors of admissible values for letters. At the same time, the hypothesis pool is large, since hypotheses had to be generated that would be able to handle seven elements, none of which had a store of specific information available, and which were scattered uniformly through several columns. The longer that processing continues without specific information, the more difficult it is to eliminate hypotheses from the pool by tagging them with unsuccessful labels. There are many sampling forays into the pool, each of which contributes to solution time and feeds information into the information pool, making it more difficult to scan at any time. The bigger the hypothesis pool, the longer the time to solution, and the lower the success rate, i.e. the smaller the $\hat{\lambda}$ value.

When the second module has been completed, the information pool is fuller than it has previously been, but this now is advantageous. A U-shaped relationship

exists between size of pool and facility in problem solving: when the information pool is full, the distance to the goal is reduced, since fewer elements remain to be solved. Through the display, the mechanism for the elimination of possibilities is evoked, which, by observing that seven out of ten letters have been assigned, and that only three digits remain unattached, and only a few columns part assigned, provides feedback to the hypothesis-generating system, indicating that only a small number of hypotheses are required to fulfill the requirements of the situation. The gap between present and desired situations is relatively small, and can be scanned in a small number of steps. The final modules for the two addition problems are identical in structure, that is, in each problem seven digits out of ten have been used, and three remain to be assigned. The $\hat{\lambda}$ values for this stage are remarkably close ($\hat{\lambda}$ =0.44 and 0.42), which lends support to the phenomenon of 'solution acceleration' explained by J.R.Hayes (1966) as being the outcome of 'planning', a process by which information is stored for use later in the solution.

This discussion concludes by postulating some of the sources of individual differences in problem-solving behaviour, which will in turn have implications for the role of memory. No mention has been made of one of the

primary souces of individual variability: the number of previously learned concepts, appropriate rules, etc. that the subject brings to the problem situation, and which cannot be divorced from the sufficiency of the hypothesis pool that he generates. The degree to which a pool contains only useful and relevant hypotheses is a function of the individual who generated that pool. The degree to which tagging of useful and useless rules occurs has already been referred to. Since tagging is related to both memory and discrimination, it is clearly variable across individuals, and will account for some of the observed differences in behaviour. A further distinction arises in the ease with which hypotheses are recalled: if each hypothesis does not have a constant or equal probability of being sampled,then additional effort will be required to sample hypotheses with lower or vacillating probabilities.

Although the arguments developed in this section are acknowledged to be both exploratory and untested, it is hoped that a tentative basis has been constructed from which to extend the theory of solving problems according to stages and modules to a level at which behaviour may be described in information-processing terms. This is not to claim, however, that the stages model provides the most adequate framework for this purpose, nor that it cannot be supplemented or superceded by other terms of reference.

7.5 The State Model

In this section, the discussion will focus on the individual letter, and the series of transformations it has to undergo before the final letter-digit assignment is made. This model cuts across divisions created by modules or individual differences, and constitutes a universal of cryptoarithmetic solutions since it can, with slight modifications, describe the state of any letter at any time.

The processing of columns appears to be an analytically useful unit of behaviour. The subject considers the column as a whole, making inferences about the letters it contains until either (a) no further progress is possible because a blind alley has been entered, or (b) a letter is converted into a digit, or the possibilities for a letter are narrowed. The subject is able to handle carries in and out of a column simultaneously with processing the letters of the column. Given that a column is to be processed, and that carries are essential operands of the processing, it is possible to list some of the more frequently encountered features of columns:

(1) d+d=l, where d is a known digit and l a letter to be assigned.

(2) $l_1+d=l_2$, where l_1, l_2 are letters to be assigned, and d is a known digit.

(3) $l_1+l_1=l_2$, where $\underline{l}_1, \underline{l}_2$ are letters to be assigned.

(4) l+l=B, where l is a letter to be assigned, and B is a binary variable.

(5) d+l=V, where d is a known digit, l is a letter to be assigned, and V is a vector of admissible values.

Observation of the situations arising in a column suggest that, given that the basic goal of cryptoarithmetic problems is to assign digits to letters, behaviour is governed by a hierarchy of subgoals leading to eventual assignments. Loosely, the subgoals are concerned with making a letter or a variable more specific, narrowing the possibilites from ten digits for each letter to something less.

We shall call the degree of specification of a letter a state. There are four states that define the quantities dealt with by the subject:

(1) l A letter about which nothing is known

(2) B A letter restricted to be even or odd

(3) V(l) A vector of admissible values for l

(4) d A digit assignment to letter l

Where (1) <(2) <(3) <4 (Eq.7.1)

In terms of increasing certainty about a letter, (4) is more specific than (3) which is more specific than (2) and so on. The goal is to increase all letters to state (4).

The subgoals are concerned with increasing the specificity of a quantity. Notice that once we have defined subgoals and their hierarchical organization, the overall specification problem by a series of subgoal achievements is essentially means-end analysis (see Chapter 1).

There are two central mechanisms for processing information: a column processor, which is concerned with increasing the state of a letter, and a change column routine, which checks the output of the column processor. More specifically, the latter routine checks whether there is conflict with another assignment, or whether the quantity processed by the column processor appears in some other column. If it finds such a column, it calls the column processor to deal with it, by generating operators. If it cannot find such a column, it will search for and enter a column which has the highest overall state value and then return control to the column processor. The processing situation can be summarized crudely in the following way, assuming that the subject has entered the column to be processed:

process column

detect features of column ——>
is it one of basic situations of Eq.7.1? NO ——> generate operator to transform it into basic situation (change column, get new goal)

YES
|
v

generate operator ——>
apply operator to column ——>
get new subgoal

The state model characterizes a system which is time-dependent, but module independent, although a module is completed when all its letters are in state (4). The model is a device for evaluating the status of any given letter at any point in the protocol, and for following the progression of any letter independently of the module structure.[8] It serves as a method for handling all the pre-module information that is circumvented and lost by a concentration on modules alone. The concept of letter-states allows a documentation of the fact that assignments, apparently of a permanent nature, can be revoked: for example, a letter that was decoded with

[8] This is related to the dynamic processing analysis presented in Table 6.4, where all possible configurations of letters within a column are ennumerated. The state model makes more specific the transitions between 1 and d in that analytic scheme.

apparently no doubt in the subject's mind, may, at some later time, perhaps even after the next module has been processed, be downgraded to a lower state, either through forgetting or through genuine doubt. The state model permits a chronological record to be made of all the states a letter has been in, up to and after it has been in state (4).

7.6 Behavioural Classificatory Scheme

More adequate analysis and classification of the large range of processes used in thinking is a *sine qua non* of a general theory of problem solving and thinking, which should include processes involved in solving several classes of problems. The General Problem Solver (GPS) does, as its name implies, aim at the development of such a broad theory. Future research should have as its objective the consolidation, rather than the proliferation, of existing classificatory schemes. With this aim in mind, some of the categories used in this section are deliberate confirmations or extensions of those developed by other workers.

The 'ideal' protocols introduced at the beginning of this chapter supplied an understanding of the gross directions and decisions made by the majority of subjects, but no discussion would be complete without a more extended analysis of the protocol material. The subject's statements

must be translated into activities, and these classified so that major, recurrent features of behaviour are isolated, couched in terms of goals.

A process takes a symbol as input, and produces a new symbol as output. A problem solver may be considered a process since he has a problem as input and, if successful, a solution as output. This definition provides the basis for dividing the observed behaviour into three categories. The first two are process categories, describing actions that operate directly on symbols, while the third is concerned with behaviour that is not integral to the problem-solving process itself, but is peripheral in function and importance. Specifically, the categories used are (I) General methods and processes, globally conceived strategies; (II) Specific, local symbol-manipulating operations and (III) Auxiliary, nonsequential information. The subcategories are mutually exclusive, but not exhaustive.

(I) General Methods and Processes

The processes described in this section operate on elements of the problem to transform them so that they represent new states of knowledge. Alternatively, they generate hypotheses and methods to test hypotheses, or are concerned with the ordering of search procedures. We will distinguish between specific and general testing processes, between references to the past and the present, and between

testing and generating hypotheses. Each category will be illustrated with the aid of quotations from the protocols.

1. GENERATING HYPOTHESES

This forms by far the largest class of processes.

(a) Hypotheses may be generated according to a very general rule

> "I should look for a completely different line and see if letters are assigned in alphabetic sequence, or something like that"

or

> "We're right through now with all our equations. Now seems a reasonable time to have a go to some intelligent guesses."

The hypotheses gathered under this heading usually refer to the use of pseudo-equations and trial and error. The use of genuine equations is less common, although it does occur in Problems 2 and 3.[9]

---

[9] Algebraic equations, as opposed to the pseudo-equations (section 7.1) are not usually a useful solution method, and it is difficult to say whether those who do not use them realize that this is true, or whether they simply do not think of using them. The few subjects, all female, who fall into this category develop a reluctance to abandon them, despite their lack of success. They appear to think that the difficulty they are experiencing stems from logical errors in the construction of the equations rather than from the inappropriate and persistent application of the method itself.

(b) Hypotheses may also be specific:

> "Lets see if we can get some idea of maximum and minimum values."

and

> "Perhaps we can further improve this inequality."

References are made to setting a letter at a particular value, and increasing or decreasing it, trying only even numbers, reversing the values inferred for two letters ("trying to get N in terms of S or viceversa").

(c) Transcending the specific-general dichotomy of hypothesis generation, there are many examples in the protocols of repeating previous methods

> "We're obliged to go through the whole procedure again."

and

(d) Rejecting a method

> "I'm going to shelve that completely. Surely we can't go through this by trial and error, can we?"

(e) Sometimes the procedure is to search for any method

> "I'll have to look around randomly for a few clues. Going to have to make a hypothesis."

2. TESTING HYPOTHESES

In this category are included some of the operations discussed in (II), the specific symbol-manipulating operations. In addition, the subject actually applies and tests the hypothesis:

> "..start in the middle, with 4 and 5, I'll go systemmatically through them now." (testing the specific hypothesis that S+D=9)

3. DIRECTION OF PROCESSING

Sometimes the subject refers to the direction in which he will process the problem, saying

> "..going the other way, to the extreme left-hand column ..."

4. DEVELOPING A COGNITIVE MAP

The subject relegates a letter or a piece of information to the future, realizing it can only be solved then:

> "Y will be the last letter to determine."

and

> "These two could be useful later on."

Sometimes reference is made to the past:

> "Well, this is going to be useful way back in the second one."

The above remarks indicate a mechanism referred to as 'remote planning' by J.R.Hayes (1966) as opposed to 'local planning' which tests the immediate consequences

of a hypothesis, by examining and becoming familiar with local terrain. This can be seen in the following example:

> "If we do that there won't be enough to carry over to make P big enough."

A related concept is that of the goal-stack assumption (Newell, 1967,p.18), where a goal that cannot be attained at present is 'pushed down' for future reference. This implies the need for some sort of 'mark' to be placed as identification on the deferred subgoal:

> "Make a note in case we have to go back to that point again."

5. REMEMBERING

Some remarks refer directly to the past, recalling a past hypothesis:

> "I was looking at that before, and decided it couldn't be simply S."

or a past assignment, or a past position or node.

6. FORGETTING

Sometimes the past is referred to because of faulty memory, usually uncertainty about a past assignment:

> "Why did I assume U was zero?"

or past hypothesis or calculation, but sometimes more generally:

> "I wish I could remember the reasons why I'd done something."

7. AWARENESS OF ERROR

This classification does not include the general suspicion that an error may have occurred, but rather an awareness not only that an error has been made, but what it was. Sometimes, though, there is less specific knowledge, so that the type of error is known, but its locus has still to be discovered:

> "Somewhere there's a relationship I missed. Try and find it."

8. STOCKTAKING

This process uses both the method of elimination and the blackboard (see III), although this is not necessarily true. It is a device to circumscribe the sphere in which operations can occur, and to assess the distance remaining to the goal. The stocktaking device provides an answer to the question "What do we know?" by saying "Let's go over what we've done." The output is expressed in a statement such as:

> "I have got both ends, but I haven't got the middle lines. Seems we got rid of both ends of the scale, 0,1,9."

9. SURVEYING AND SCANNING

A less frequently used strategy, and one that tends to be evoked in times of difficulty, is that of surveying the problem for further information or implication. The process may also employ the method of elimination to

restrict the domain of operation:

> "We survey the problem again with the new bit of information, and see if there are any other letters we can eliminate."

Under this category is included the preliminary scanning that occurs when the problem is first presented, and where the output consists of an ennumeration of the number of letters in the problem, or the realization that the word 'HANNAH' is "reversible, it's halfway there" or that the word 'PEEP' is a palindrome. It differs from III.3 because it produces output.

10. RETURN TO AN EARLIER NODE

This process is not to be confused with I.6 where the subject searches his memory to find a reason for a previous action. The process of returning to an earlier node occurs when a line of reasoning proves impossible. The subject then either

(a) backtracks: search backs up to some specified earlier point:

"So we'll have to go back to the NRB stage."

or

(b) returns to the beginning.

11. ERRORS

Bartlett (1968) points out that preliminary allocation of digits to letters may be viewed as an

uncritically accepted generalization, which, once made, may prevent further questioning. This is certainly one type of error that is frequently encountered. The subject makes a tentative assignment, not based on any clear reasoning, which he then treats as final, forgetting the provisional nature of the initial assignment.

In many cases, the creation of an erroneous relationship is handled by the same rules that govern the generation of correct hypotheses or correct testing. However one may distinguish two kinds of errors (Donaldson, 1963):

(a) executive, which arise from misreading of letters or columns, or computational mistakes, and

(b) structural, where a carry is wrongly inferred, or a relationship falsely stated.

There is also evidence in the protocols to support two of the 'check-list' variables mentioned by Kilpatrick (1967,p.56-57). The first is that of exploratory manipulation, to test if some useful fact might be discovered, but not a heuristic to test a specific solution. The second is that of successive approximation, where information from an unsuccessful trial is used to set the value for a subsequent trial.

A further process that appears in the multiplication problem is that of list searching(J.R.Hayes, 1965) This is contrasted to direct accessing which is a means of gaining access to a particular entry on a list, rather than searching through the entire list. There are no examples of the latter mechanism in the protocols, which is probably found more frequently in rote or paired associate learning.(However, the method is used to retrieve information from an external representation - see p.287). Situations in which list searching occurs are those where the subject searches through all the perfect squares or for all products of two numbers that have a certain digit in the units position. Presumably the subject is searching through a subset of the multiplication tableau of which he has an internal representation.

The "planning" heuristic referred to in Chapter 4 does not occur to any significant extent in any of the three problems apart from the example mentioned earlier (p. 234).

(II) Specific Symbol-Manipulating Processes

We have seen that the problem solver uses a large variety of methods and hypotheses during the processing of a problem. The purpose of this section is to focus closer and observe the minutiae of processing as it involves the actual elements of the problem space. Quotations from protocols will not be used unless this becomes necessary in the interest of clarity.

The following actions occur:

1. Digits are assigned to letters.
2. Vectors of digits are generated according to certain rules (even-ness, greater than 5 etc.).
3. Relationships or values are ascertained through inference. It is usually clear from the protocol when a digit is tentatively inferred or definitely assigned (see section 4.4).
4. Values are eliminated from the admissible range, either because they have been previously assigned, or because they are logically impossible.
5. The subject is able to reduce the range of values he is considering by conceptualizing relationships between letters and digits, or between letters, in terms of:
   (a) Equality (not assignment): "E must be 9."
   (b) Inequality: "L must be greater than 5."
   (c) Binary properties: "R must be odd."
   (d) Dichotomies: "L is either 3 or 8."
   (e) Vectors: "R must be 5,7 or 9." (This is more specific than (c) since it says "R odd, and $R \geq 5$."
   (f) Specific properties, such as primes, or "divisible by mod 11"
   (g) Arbitrariness: a letter can take on any value, it is "free."

(h) Maximum and minimum values: "$5 \leq A+S \geq 16$"

(i) Vague limits, such as "something high" or "the very worst this can be".

6. The subject transforms information into equation form, usually quasi-equational, since the statement simply expresses the structure of a column, as for example "N+M+S=T", although genuine algebraic transformations do occur.

7. The subject manipulates the carries to and from a column.

   (a) They may be inferred : "A has to be greater than 3, which means you can carry 1."

   (b) They may be the objects of search: "Can you carry from over there? Must be a carry over there, mustn't there?"

   (c) They are assigned.

   (d) Attempts are made to limit their range of admissibility: "These carries don't always have to be 1, that's the trouble."

   (e) They are used to make inferences about a letter: "H must be 9, because of the carry of 1".

8. The units of processing may be

   (a) atomic, that is, individual letters or carries.

   (b) columns

   (c) rows. This is less common, and occurs mainly in

Problem 2, where both product and multiplicand are treated as units but sometimes also in the isolating of the word 'HANNAH' in Problem 3.

(d) parts of rows: "The sum of NNAH,MUST and TUDY.. the sum of these three must be greater than 3,000, sorry, greater than 30,000."

(e) the whole problem. This is rare, but occurs when the multiplication problem is treated as a palindrome, or transformed into a division.

(III) Auxiliary and Extraneous Behaviour

The variables in this category consist of actions not essential to the solution process, since they do not aid directly in the processing of information, and are, for that reason, outside the input-output paradigms. Functions served by these variables are those of orienting, commenting, organizing. They often occur when the subject is not making progress, when they seem to be evoked in order to fill in a period of silence.

1. MENTIONING RULES OR MAKING RULES EXPLICIT

Reference is made to one of the three rules for the problem, in the form of remarks like:

> "Using condition 2, that different letters stand for different digits...."

and

> "This uniqueness makes a whole big difference."

2. CHECKING

This action is performed entirely to check on previous operations and calculations. No new information is produced, and the action has heuristic value only if the outcome of the checking reveals that an operation or calculation was wrong. Otherwise it is simply time consuming. It differs from the stocktaking process (I.8) since the latter actually consolidates information for the purpose of extracting new information. Examples of statements using this checking device are:

> "Let's just go over what we've done, I want to check my arithmetic again."

and, more specifically,

> "I want to check we haven't used the same for any numbers."

3. GENERAL REFERENCE TO PROBLEM

This is a frequently used mechanism, but again does not produce any new information. The subject simply comments on the problem, its difficulty, its similarity to other problems etc. A subject says:

> "I have come to the conclusion that this problem isn't easily solved. It's a lot harder than the last one."

Another subject, familiar with cryptoarithmetic problems, remarks:

> "Usually you get more useful solutions from both ends rather than from the middle of one of these."

4. EVALUATION

There are two types of evaluative statements. Again, their function is not to produce new information, nor to generate new hypotheses. Instead they function to assess or evaluate without prescribing any action.

(a) Evaluation of position

A subject may evaluate his present position in a general way,

> "We seem to be losing ground at this point and not getting many fruitful possibilities."

or in a more specific way,

> "We've got E, that's the biggest hurdle over."

He may also evaluate a previous position, or a past hypothesis:

> "Probably unfortunate starting in the lower numbers. That didn't seem hopeful before."

(b) Evaluation of method

This evaluative function is to be distinguished from generating or applying a method as described in I.1 and I.2. The examples will clarify the difference:

> "That's not a safe assumption, but it's a useful one for the moment."

and

> "This is a bit awkward. Still it gets us a bit nearer for getting a value for S."

6. BLACKBOARD

An extremely important device is the construction of a blackboard or visible memory which externalizes information, aiding in the handling and retrieval of past processing. The tableau displays described previously are a manifestation of this device, and not identical with it. In the same way, the stocktaking process (I.8) uses the blackboard device, but is distinct from it. The blackboard provides a second problem space; it functions not only to display accumulated information but also to simply record all the digits and letters occurring in the problem. For example, a subject says

> "I must make some methodical start by putting down different letters so that I can write the values assigned to them as I come to them."

In this situation, the subject is producing externally represented lists which will facilitate search. The above example refers to the construction of a blackboard, while in other statements a subject can be seen actually using it:

> "So that eliminates L to be, if we look back somewhere we'll find it, if we can find it on the paper."

This is an example of the direct accessing retrieval method without which problem solving would be more difficult. Nevertheless, the blackboard device is not

a process, but is itself used by certain processes.

7. QUESTIONING EXISTENCE/UNIQUENESS OF SOLUTION

This variable has been noted in previous research (Kilpatrick,1967). It is clear that this type of statement usually occurs when the subject is confronted by obstacles to further progress. However, it also appears, especially in the multiplication solutions, once the answer has been obtained. For example,

> "I suspect there's more than one answer to it."

or, in more specific terms,

> "That's one solution, and any other solution must have I greater than or equal to 6."

This chapter has been both exploratory and impressionistic in its interpretation of the protocol data. Hopefully, some compromise has been reached between retaining idiosyncratic features of behaviour and describing patterns that transcend individual differences. Explanations have been offered for the effects of verbalization, and a model developed to extend the concepts of stages and modules where they can be handled by information-processing language. One of the objectives of this thesis will be achieved if the 'thinking aloud' technique and the analysis of the resultant protocols can be shown to be both a fruitful source of data and hypotheses, and an encouragement to further research.

CHAPTER 8

SUMMARY AND CONCLUSIONS

Surveyors of the psychological landscape have questioned whether problem solving constitutes its own domain, or whether it is more properly placed under the sovereignty of learning theory (Duncan,1959; G.Davis, 1966; Kleinmuntz,1966). Whether problem solving differs in kind, or only in degree from the activities classed as learning is an issue mainly between behaviourism and information-processing theory, although it is clear that other protagonists such as the Gestalt school, would be eligible to enter the fray.

The behaviourist contention is that problem solving differs from learning only in the degree to which it requires the discovery of a new response or the integration of previously learned responses (Duncan, 1959), or in the degree to which the response alternatives are not clearly defined for the subject (G.Davis, 1966): it is natural therefore that the unit of analysis for this approach should be the S-R unit, and the explanatory laws those of learning theory. One behaviourist comment is offered by Skinner (1966):

> Problem-solving is concerned with the relations which prevail among three terms: a stimulus, a response, and a reinforcing consequence...
>
> When a response occurs and is reinforced, the probability that it will occur again in the presence of similar stimuli is increased... Problems arise when contingencies are complex. For example, there may be no response available which satisfies a given set of contingencies. (p.226)

Although Skinner concedes that problem solving is "a behavioural event", he adds that this "does not however, necessarily reflect an important behavioural process." (p.240)

Information-processing theorists on the other hand, accept the legitimacy of complex events, and stress the need for a clear understanding of the way in which problems are internally represented by the problem solver. An information-processing theory of cognitive processes would then, disavow the assumption that these complex internal events either can, or should, be described by the laws of operant and classical conditioning.

The difference between the two views we have just outlined may not be as great as it first appears: both theories insist on an elementalistic, rather than a holistic, approach, but information-processing theory emphasizes from the beginning the ways in which these elements or primitive information processes, are organized into hierarchical structures. The behaviourist, on the

other hand, must invoke heirarchies, mediational mechanisms and internal reinforcement to account for behaviour which cannot be explained in any other way, but this attempt is often unconvincing, lacking the flavour of the behaviour it was intended to explain.

This divergence of opinion leads naturally into a further source of disagreement: whether internal events or external events should be emphasized when problem-solving behaviour is being described. The behaviourists stress the external conditions, attempting to explain problem solving as the acquisition of rules through complex reinforcement contingencies (Staats,1966). Preoccupations with the nature of these rules, and with the internal events that accompany their acquisition, are dismissed as unscientific. This is easily countered however, by maintaining that information-processing theory is simply a codification of input-output transformation rules, a codification of the observed relationships between stimulus and response (Green, 1966).

Closely related to these differences in emphasis is the choice of experimental paradigm to be used. The behaviourist designs experiments which will control the processes of unsophisticated subjects, often animals, and his blueprint describes the environmental manipulations

by which he does this. In contrast, the information-processing theorist tries to infer the cognitive processes of sophisticated subjects, and he selects his subjects and the problems he will use in accordance with this. He aims for a model or a program which contains precise and rigorous specifications of these inferences, and contends that problem solving has been adequately described when a set of processes is sufficient to solve a particular problem. Skinner however, takes issue with these aims and methods, advocating a rejection of all considerations of inner activities such as the processing, retrieving and storing of information which, he says, take place "out of sight, if not out of mind" (1966, p.256).

It may be that these two approaches are complementary rather than antipathetic: a detailing understanding of the internal processes and a precise account of the external events are both necessary for a complete description of the problem solver. Unfortunately, it appears that the opposition of these two schools of thought precludes, for the moment, their investigating compatible experimental situations and organisms, so that the current prospect for their integration is discouraging. At present, we are left with something less than an uneasy truce, with the

behaviourist condemning concepts of internal processing as dangerous and the information-processing theorist regarding the quest for S-R associations as unproductive.

Rather than attempt a reconciliation between these views, or definitive statements concerning the rightful place of problem solving in psychology, we conclude this discussion with the contention that the present state of psychology is not such as to warrant the search for comprehensive definitions of either learning or problem solving, or the rejection of any global approaches to problem solving that appear to be leading to fruitful experimentation and model building. Instead of seeking umbrella terms, it is wiser to assert that, however these fields are delineated, they continue to denote distinct empirical domains. The view expressed in this thesis is that it is profitable to regard problem solving as though it were a complex, integrated, multidimensional sphere, which is unitary in nature despite the diversity of tasks to which it may be directed. This is not to deny that other fundamental processes, such as learning or perception, can be ignored in any theory of problem solving; on the other hand, there is little to be gained from a reductionist approach in an area where some larger vision may more satisfactorily capture the special features of the behaviour under investigation. No apology is made for this bias,

one which, on the part of other investigators, appears to have been responsible for substantial contributions to our understanding of the way people solve problems.

This thesis has wandered down more corridors than it has been able to explore thoroughly, but some of its objectives will have been fulfilled if it has pointed to hypotheses which may be confirmed or expanded by future research. The attempt to integrate information-processing concepts with a real time statistical model is, as far as is known, unique. It is hoped that this juxtaposition has contributed to the enrichment of both models. We shall be able to assess this claim more thoroughly after considering the specific conclusions that may be drawn from the results of the study.

One of the features of this research has been a concern with one type of external event, namely that inherent in the structure of the problem itself. It is only in a thorough analysis of the structure of a problem that the processes of the problem solver can be properly appreciated. In addition to this interest in the problem itself, an attempt has been made to integrate the several dimensions and facets of the problem solver by considering him as part of a problem-solving system. Given a particular problem, the solver was investigated

on several levels. First, his response to that problem was documented by the time he required to solve it, and this was compared to the times he required to solve problems of different complexity. Next, we considered the shape of the tree structures he generated while searching for the solution. Finally, we attempted to retrieve the program he constructed to solve that problem, that is, the processes and methods that could be inferred from the protocols he produced while working towards the problem solution.

This study also set out specifically to test the assumptions made by the stages model developed by Restle and Davis (1962). Since further testing was impossible unless the individual stages could be excavated and held to the light of statistical validation, it was necessary to adopt the 'thinking aloud' method currently undergoing a renaissance with computer simulation of cognitive processes. It was hypothesized that clues about the structure and behaviour of these stages would be contained in the protocols produced by this method. A useful by-product of using the 'thinking aloud' technique was new evidence concerning the effects of verbalization on problem-solving behaviour. The results of the study indicated that verbalization, although it has complex effects on individuals, does not act to attenuate or

decrease the solution times for groups of individuals. This is consonant with most of the scanty evidence available on the subject (Hafner, 1957; Dansereau and Gregg,1966).

Before the stages model could be subjected to this experimental verification however, it was essential to clarify what was meant by a stage, a concept which had been ambiguously handled in the previous literature (Restle and Davis, 1962; Davis and Restle, 1963; J.H.Davis,1964) Clarification of the concept necessitated a two-fold definition in which stage referred to the statistical and temporal aspects of the components of the solution process, and module referred to the components of the problem structure. The two concepts were related by the formula that a stage describes the time taken to solve a module. The preciseness of this definition suggests that the stages-module model may only be appropriate for a restricted class of problems. It would be important to establish, through future experimentation, the class of problems to which the model may be applied. It could be suitable, for example, as a description of the solution processes of such apparently modular problems as crossword puzzles, cryptography, or certain types of mechanical puzzles. Until the class of strictly sequentially structured problems has been investigated, it may be

opportune to postpone the discussion of problems with 'parts', that is,those in which the structural components may be solved in any order.

By using maximum likelihood estimators instead of the method of moments to estimate the number of stages in the solution process, it was found that the number of stages (and, therefore, the number of modules) estimated from the solution times data did not accord with the number of modules isolated through protocol analysis. This was true even though the distribution of times to solve each module had the requisite statistical properties of a stage. If the component modules had been equally difficult, and if, by implication, the stages had each been characterized by the same values of mean time and $\hat{\lambda}$, it is likely that the $\hat{k}$ estimated from the solution times would have more accurately reflected the 'true' protocol-based K. This shortcoming could be remedied in two ways through future experimentation: a more realistic version of the gamma distribution, such as the generalized gamma distribution proposed by McGill (1963), could be used to model solution times, or some appropriate combination of birth and death processes, in which the time constants are allowed to increase or decrease respectively, could be developed in continuous form (Feller 1960,p.408). Alternatively, the

remedy could lie in testing the original version of the gamma model with problems that have been constructed with the objective of homogeneous modularity (Hayes, 1966). With appropriate experimental designs, these alternatives could be explored as candidates for a more accurate descriptive model.

Despite the fact that the model did not adequately reflect the component processes of the solution process, the gamma distribution provided an unfailingly satisfactory fit to the empirical distribution of solution times; it was, however, known that the assumptions of both independence and equivalence of the subprocesses had been violated. This leads to a few cautionary words about the concept of goodness-of-fit. The merit of a mathematical model lies not so much in describing the data as in its accuracy, and perhaps predictive value, relative to other mathematical models. It is elementary that if models have a sufficient number of parameters, then values for these parameters can always be found that will allow a curve to pass through the data points. This can be achieved by several models for any given experiment and it is among these that the search for an optimal model should proceed, using as criteria the ability of the model to predict the fine structure or the sequential details of the data (Bush and

Mosteller, 1959). The maxim that goodness-of-fit should be measured, not tested, is highlighted by the fact that the gamma distribution, as evaluated by goodness-of-fit tests, is clearly insensitive to violations that are known to have occurred.

It is perhaps axiomatic in science that as a model tends to describe more and more it may in fact be explaining less and less in any particular application. We have seen that the gamma model represents many of the phenomena to be found both in psychology and other scientfic areas, and that it seems to be a fair description of the time to solution of cryptoarithmetic problems. However, the fit of the gamma distribution to these data does not give us any direct insight into the process of problem solving itself; in other words, the distribution suffices as description, but cannot be regarded to be of any explanatoryvalue in problem-solving behaviour. This is, of course, no more than we would expect of the gamma distribution, or any purely quantitative manipulation of complex behavioural data.

It would be inconvenient, if not impossible, to test by empirical methods the effects of deliberately violating the assumptions made by the stages-module model. By a series of Monte Carlo experiments however, it was

possible to simulate conditions in which the λ values for either subjects or stages, or both, varied according to different hypotheses. Both the size and direction of these violations could be manipulated, so that the accuracy with which these different versions of the model estimated the parameter values of the population could be tested. Briefly, the outcome of the Monte Carlo experiments indicated that the original version of the model, in which the scale factor (λ) is constant for all subjects on all stages, yielded, for most of the values that were tested, a less accurate estimate of the population parameter (K) than did the version in which the scale factor was systemmatically varied for the various stages but kept constant for the subjects. If the number of modules is an index of complexity or difficulty of a problem, then it is a less satisfactory index than others that are available. The range of $\hat{k}$ values estimated for the three problems used in the study was not sufficiently large to discriminate between them, but this can not be attributed to the vagaries of the statistical methods used, or to the size of the samples involved. Presumably the number of modules isolated from the protocols represent a 'real' index of the problem structure, yet this number was, at best, the same for two problems which, by other measures, differed considerably

in difficulty. A far more sensitive index was provided by a combination of length of solution time and $\hat{\lambda}$ for each problem; that is, as mean time, and therefore presumed difficulty, increased, the value of $\hat{\lambda}$ decreased. This reinforces the intuitively satisfying proposition that the more difficult a problem is, the lower is the probability rate for its solution. An idealized model would predict that as the value of K increased, so would solution time, while λ would decrease. It is conceivable that this situation could be achieved, and this possibility would be maximized by using problems with homogeneous modularity in the hope that the estimated $\hat{k}$ would closely reflect the actual K. It may also be possible to achieve closer estimates of the structure of the problem through controlling the search time between modules, which appears to have been a major source of variance, by presenting the subjects with an outline of the modules or solution paths of the problem.

In a latter part of the thesis, an attempt was made to extend the stages-module model to describe the processes exhibited in solving modular problems. We asked if the model provided a template to be placed over specific parts of the solution process to capture recurrences of behaviour that appeared at, say, the beginning

or the middle, or the end of a module. Although the answer to this question was negative, the model proved useful for distinguishing phases of the solution process, that is, for pointing to the division of processing that occurred between and within modules. A highly speculative theory was outlined through which information processes could be tied to an abstracted version of the module structure of cryptoarithmetic problems. This involved the concept of hypothesis pools which are generated in accordance with the individual requirements of a module, the anticipation of a module being heralded by evoking a representation of the problem by means of the 'display' routine.

The inferred processes and methods of problem solving were classified, and it was hoped that the resulting scheme was useful both as a potential classificatory scheme for other problems, and as offering substantiation for mechanisms reported by other workers.

What, by way of summary, can we say about the stages-modules model? Compared to the stages models which abound in psychology textbooks, and which are aimed at a vague and general description of functional segments of the problem-solving process, independently of the solver or the problem used, the model presented in this thesis

is very specific and highly defined. It does not provide a complete description of problem-solving behaviour,but focusses instead on only a small part of it, that is, the complexity of a problem, and therefore of the problem-solving process, whose overall difficulty is reflected in the probability rate (λ) manifested by the individuals who have solved it. It does not purport to be a detailed model of the sequential aspects of behaviour; indeed this information has to be found in other ways, such as protocol analysis, but may be used to supplement the model. The clear need at this point of our knowledge of problem solving is to develop models of the middle-range (Merton,1966) which accommodate individual differences not described by the stages of the gamma function or the modules of the problem structure, but which are less phenomenological than the fine structure of behaviour incorporated in a protocol.

By distinguishing between module and stage, we can elaborate to person and problem, and by integrating these into one system, the problem-solving system, the potential of the model may be extended to explanation and, perhaps, prediction.

We begin this elaboration by considering that the problem structure, as reflected in the parameter K, is immutable, and transcends individual differences. However, the problem-solving system must specify individual dimensions, such as, for example, age and experience, which appear to be critical in the determination of problem-solving behaviour. These individual differences are expressed as differences in λ.

We now assume an edifice constructed of two parallel structures, one of which is labelled $\underline{K}$, and the other labelled $\underline{\lambda}$. K and λ are related in the following way:

$$\bar{X}=\frac{K}{\lambda}$$

K, being an immutable property of the problem, is a constant, independent of which person is solving it. Individual differences are therefore manifested in the values of the mean time to solution ($\bar{X}$) and λ. The structure labelled K represents the universe of problems, the structure labelled λ represents the universe of solvers, and between them, labelled $\bar{X}$, is the experimental situation which measures solution times. (The experiments in this study indicate a wider range of solvers than problems, but this need not concern us at present). Consider that we have, in the spirit of the objectives expressed earlier in Chapter 6, found an optimal

match between these two systems, so that their interaction results in a vigorous experimental situation.

Within the universe of solvers, then, some individuals are more competent than others, and within the universe of problems, some problems are more complex than others. For a given solver, the problems he can solve are all those at or below his position in the parallel structure. Because he can solve a wider range of problems, the distance between successive problems is smaller for the competent problem solver. For a less able solver, whose capacity extends over a narrower range, the distance between successive problems is larger. The distance between problems increase as the problems increase in complexity.

This is related to λ in the following way. λ reflects an individual's probability rate of solution, so that the easier a problem is, the larger is the resulting value of λ, i.e. the greater is the probability of solving that problem. The size of λ is decreased or increased by a smaller fraction for the competent than for the less competent solver as the complexity of the problem increases or decreases. Since the value of K is immutable, the competent solver will have a lower solution time and a higher λ rate for a given value of K than will the less competent solver who, if that value of K is to remain constant, will have to exhibit a

a higher solution time and a lower λ rate. This implies that, given a certain value of K, the difficulty of a problem will be reflected in the values of an individual's solution time and solution rate (λ).

The findings reported in this study indicate that, as a problem increases in difficulty, variance increases. This suggests that the λ values of competent solvers are affected differently from those of less competent solvers: these values decrease or increase at a lesser rate proportionally to the value of K, since they represent a smaller proportion of the relevent problem universe.

We may assume that we have established a typology of problems arranged along a dimension of K values (developed with the aid of the suggestions and cautions outlined above). The time taken by an individual to solve a particular problem will indicate his λ value for that problem ($\lambda=K/\bar{X}$). The significance of this information is extended by investigating an individual on a range of problems, or by comparing individuals on the same problem, and thus providing a test for the hypotheses we have developed to explain the relationship between individual differences and problem complexity. In the absence of independent measures of K, it has not been possible to determine the value of λ for an individual;

instead, under the assumption that this value was constant for all individuals on a given problem, λ was derived from the total solution times data. Knowledge of individual scale factors, perhaps even the construction of a typology of individuals on the λ dimension, parallel to the typology of problems constructed on the K dimension, would provide invaluable information above the values of the time constants to be used in constructing the birth-death process or the generalized gamma model referred to earlier.

Our approach has been to probe deeply into a small area of problem solving rather than to cast a less penetrating net over a wider area. In doing so, we have treated problem solving as a complex process, viewing the problem in terms of its modular and elemental structure, the solver in terms of his information-processing capacities and his temporal performance as measured by the stages model, and the interaction of problem and solver in solution paths. Finally, the parallel structures of stages and modules were integrated into a problem-solving system to account for the individual differences that exist in spite of the immutability of the problem structure. It is hoped that this thesis has shown that problem solving, when inspected through a variety of lenses, displays to the observer a multidimensional edifice which, for most purposes, may be considered as growing out of a unitary foundation.

APPENDIX A

Example of problem sheet used in experiments

D O N A L D

G E R A L D

R O B E R T

This is an addition problem in which you are given that

D = 5.

Remember: (1) Letters stand for digits between 0 and 9.

(2) Different letters stand for different digits.

(3) Zero cannot be assigned to the left hand digit in any row.

REMEMBER TO SAY 'I HAVE FINISHED' WHEN YOU ARE SATISFIED WITH YOUR SOLUTION.

APPENDIX B

Table I: Mean solution times for experienced and inexperienced groups.

Table II: Mean solution times for adult and student groups.

TABLE I

Mean solution times (minutes) and variance for Problem 1 (DONALD) and Problem 3 (HANNAH) according to previous experience with crypto-arithmetic problems

Problem	Condition	Experienced			Inexperienced			t	F
		N	Mean time	Variance	N	Mean time	Variance		
1 (DONALD)	Verbal	21	21.34	130.60	20	31.47	381.15	2.01	2.92*
	Nonverbal	13	16.03	86.87	40	30.54	167.80	4.40**	1.93
	Verbal and Nonverbal	34	19.33	117.41	60	30.85	233.86	4.25**	1.99*
3 (HANNAH)	Verbal	16	35.17	287.68	13	33.42	268.55	0.28	1.07
	Nonverbal	14	30.48	299.26	24	40.51	225.83	1.81	1.32
	Verbal and Nonverbal	30	32.97	278.99	37	38.02	245.55	1.28	1.14

* $p<.05$

** $p<.01$

TABLE II

Mean solution times (minutes) and variance for Problem 1 (DONALD) and Problem 3 (HANNAH) according to age groups

Problem	Condition	Adults			Students			t	F
		N	Mean	Variance	N	Mean	Variance		
1 (DONALD)	Verbal	23	16.34	63.08	18	38.99	254.72	5.51**	4.04**
	Nonverbal	16	18.53	102.08	37	30.56	179.72	3.59**	1.76
	Verbal and Nonverbal	39	17.24	78.06	55	33.24	212.43	6.61**	2.72**
3 (HANNAH)	Verbal	20	30.62	220.26	9	42.18	290.46	1.76	1.32
	Nonverbal	21	31.39	315.18	17	41.75	181.47	2.05*	1.74
	Verbal and Nonverbal	41	31.01	262.31	26	41.90	209.31	2.87**	1.25

* $p<.05$

** $p<.01$

APPENDIX C

Table I: Points of maximum difference between empirical and gamma distributions in Figures 4.1 to 4.3.

Table II: Points of maximum difference between empirical and exponential distributions in Figures 4.4 and 4.5.

TABLE I

Figure	Problem	Group	Size of maximum difference (D)*	Point of maximum difference (minutes)
4.1	1 (DONALD)	Verbal	.1167	21.7666
		Nonverbal	.1062	9.9166
		Verbal & Nonverbal	.0788	39.0000
4.2	2 (SIR)	Verbal	.1633	28.0333
		Nonverbal	.1576	27.2333
		Verbal & Nonverbal	.1481	27.2333
4.3	3 (HANNAH)	Verbal	.1199	45.1833
		Nonverbal	.1949	43.0833
		Verbal & Nonverbal	.1490	43.0833

* D = maximum $F_o(X)-S_N(X)$, where F = theoretical distribution and S = empiricial distribution.

TABLE II

Figure	Problem	Stage	Size of maximum difference (D)*	Point of maximum difference (minutes)
4.4	1	1	.1427	0.1500
	(DONALD)	2	.0723	2.3000
		3	0.620	2.2833
		4	.0885	0.6333
4.5	3	1	.0856	1.5500
	(HANNAH)	2	.0405	4.7333
		3	.0887	0.8166

* D = maximum $F_o(X)-S_N(X)$, where F = theoretical distribution and S = empirical distribution.

APPENDIX D

Table I: Table of mean and standard error of sampling distribution of $\hat{k}$ for K = 1-9 for several sample sizes.

Table II: Table of 95% confidence intervals for mean of sampling distribution of $\hat{k}$ and $\hat{\lambda}$ (K = 1-7)

Graph showing sampling distribution of $\hat{k}$ for K = 1-9 for N = 30 with $\lambda = 0.10$ (solid line) and $\lambda = 0.30$ (dotted line)

Graph showing standard error of $\hat{k}$ and $\hat{\lambda}$.

TABLE I

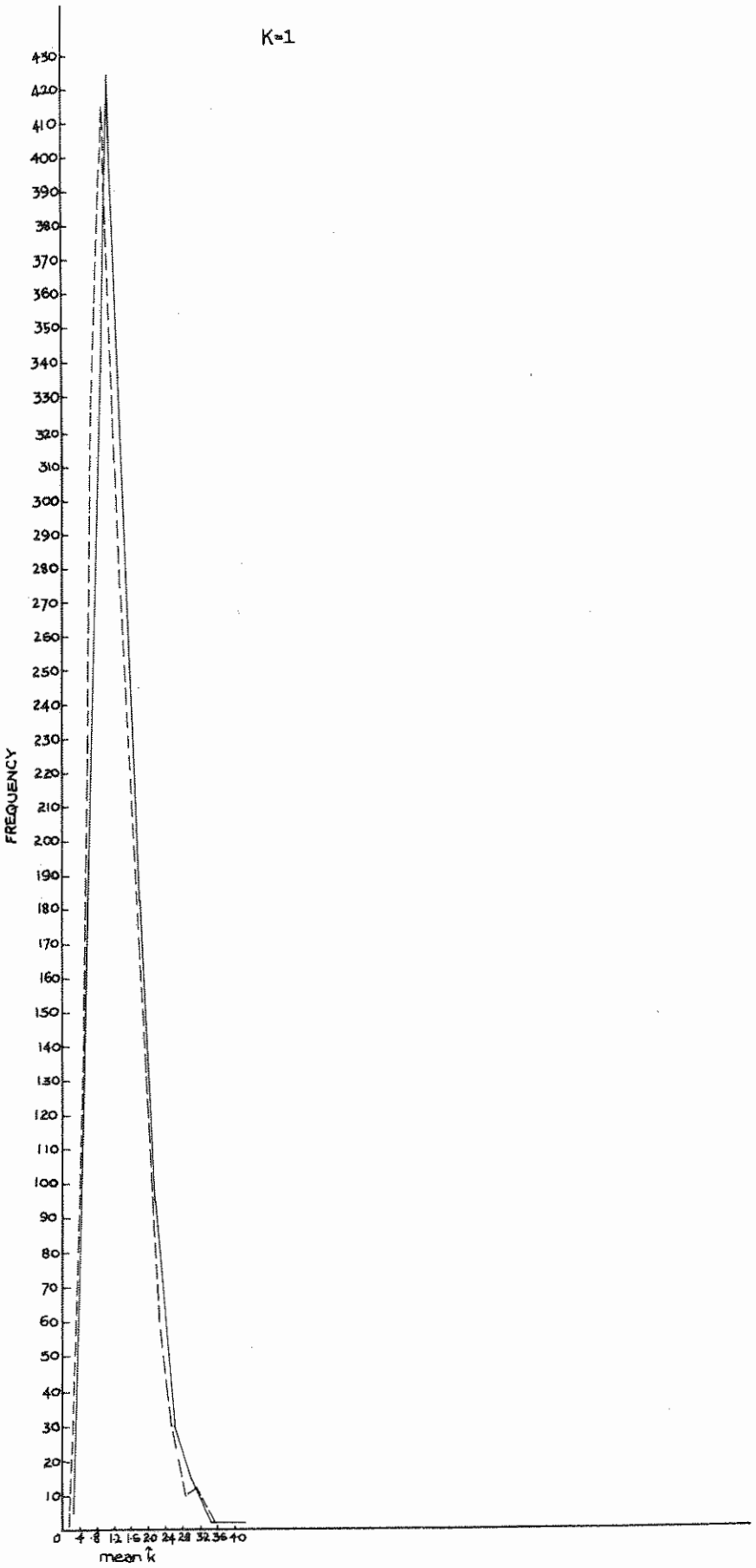
Mean and standard error (S.E.) of sampling distribution of $\hat{k}$ for several sample sizes

K	N	Mean $\hat{k}$	S.E.
1	15	1.3047	0.5710
	30	1.1394	0.3621
	40	1.1156	0.3309
	50	1.0865	0.2702
	60	1.0724	0.2544
	70	1.0716	0.2269
	95	1.0410	0.2048
2	15	2.4440	1.0607
	40	2.1517	0.5662
3	15	3.5629	1.6915
	30	3.2629	1.0267
	40	3.2243	0.7974
	50	3.1495	0.6715
	70	3.1392	0.5796
	95	3.1045	0.5016
4	15	4.8947	2.3121
	30	4.4123	1.2828
	35	4.2924	1.1681
	40	4.2416	1.0073
	50	4.1983	0.8516
	70	4.1602	0.8074
	95	4.1361	0.6542
5	15	6.0822	3.2656
	30	5.3810	1.5706
	40	5.3260	1.2853
	70	5.2148	0.9689
6	15	7.0480	3.0605
	30	6.4774	1.9741
7	15	8.3489	4.1712
	18	7.8616	3.0166
8	15	9.2897	3.8621
9	15	10.7288	4.6607

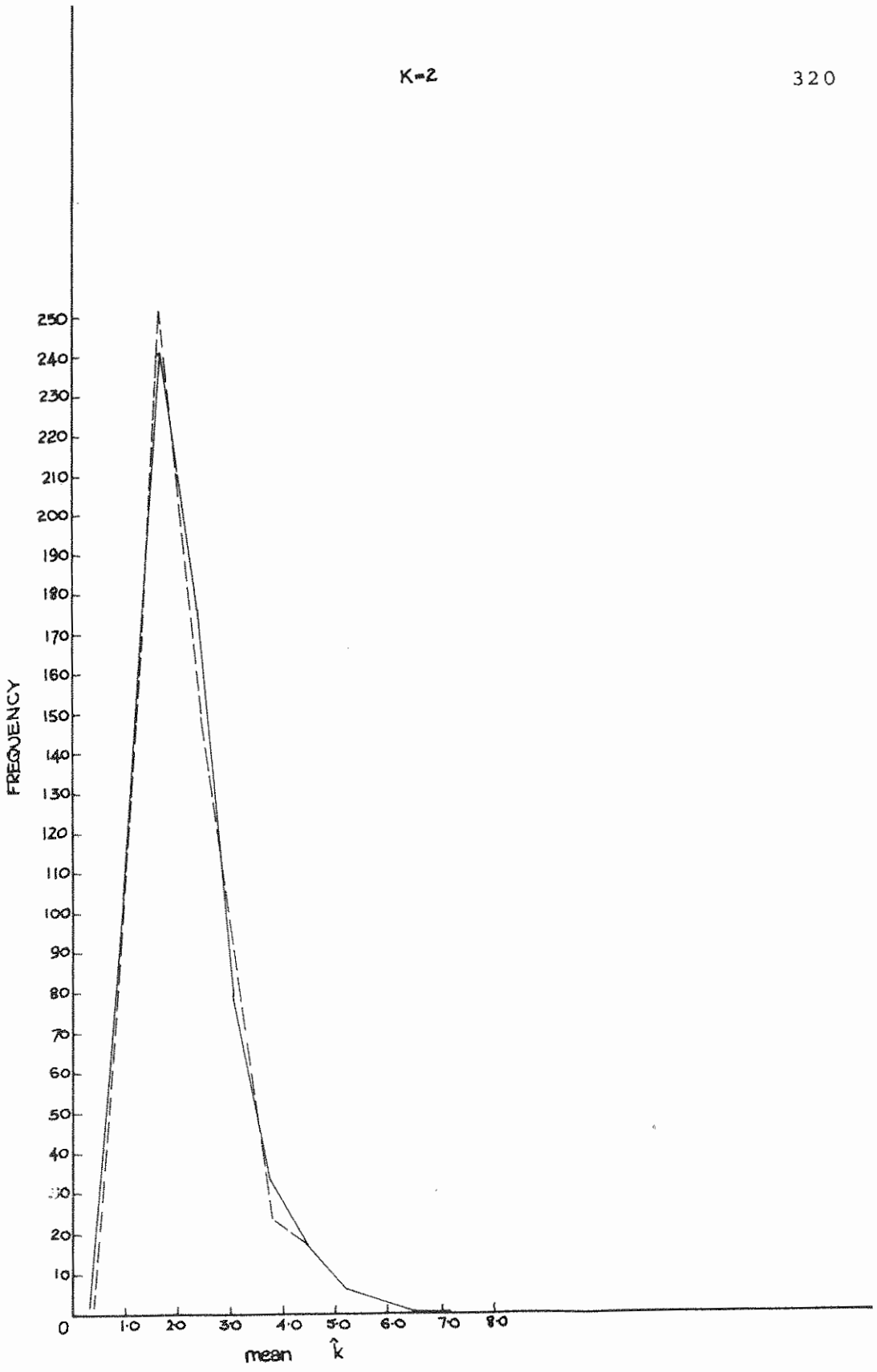
TABLE II

95% confidence intervals for mean of sampling distribution of $\hat{k}$ and $\hat{\lambda}$

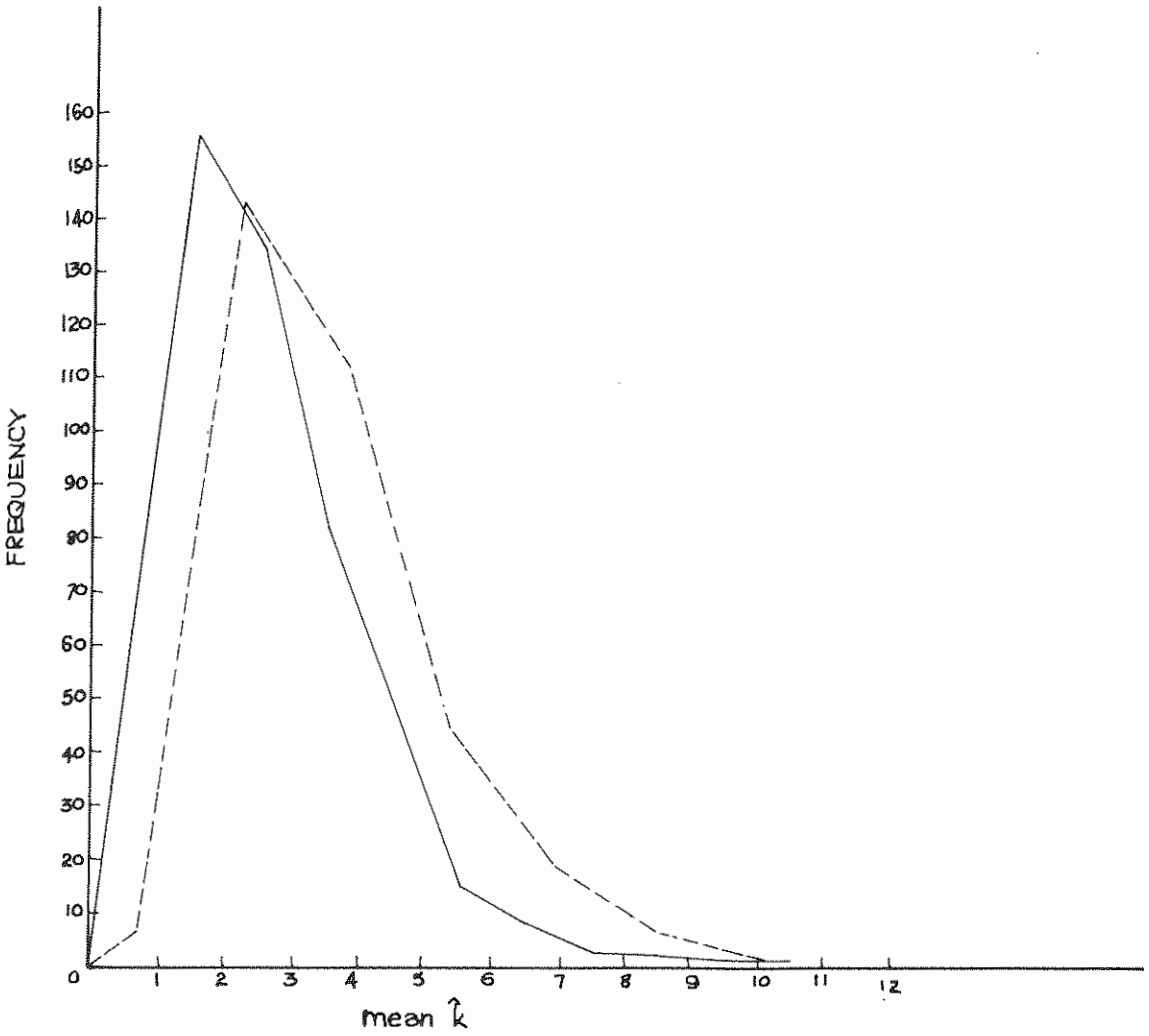
K	N	95% confidence interval ($\hat{k}$)		95% confidence interval ($\hat{\lambda}$)	
1	15	0.4152	2.6000	0.0338	0.3420
	29	0.5375	2.1250	0.0527	0.2400
	40	0.5400	1.7800	0.0480	0.2040
2	15	0.8026	5.1930	0.0367	0.2825
	40	1.1000	3.3000	0.0547	0.1800
3	15	1.2030	7.8500	0.0395	0.2940
	29	1.6005	5.5840	0.0540	0.1983
	38	1.8230	5.1250	0.0581	0.1725
	41	1.8216	5.1000	0.0606	0.1760
	53	1.8962	4.6333	0.0632	0.1610
	67	2.0831	4.5000	0.0671	0.1526
	95	2.1430	4.2562	0.0691	0.1434
4	15	1.6085	11.5567	0.0386	0.2967
	29	2.2500	6.8000	0.0550	0.1792
	34	2.2764	7.0500	0.0572	0.1850
	41	2.4275	6.5700	0.0580	0.1700
	53	2.6364	6.0899	0.0636	0.1575
	67	2.6754	5.9150	0.0660	0.1543
	94	2.8500	5.5889	0.0712	0.1425
5	15	1.7000	16.2500	0.0384	0.2983
	38	2.8500	9.0000	0.0570	0.1820
	67	3.3300	7.4400	0.0658	0.1505
6	15	2.2851	14.2500	0.0467	0.2488
7	15	1.8188	18.5625	0.0264	0.2625

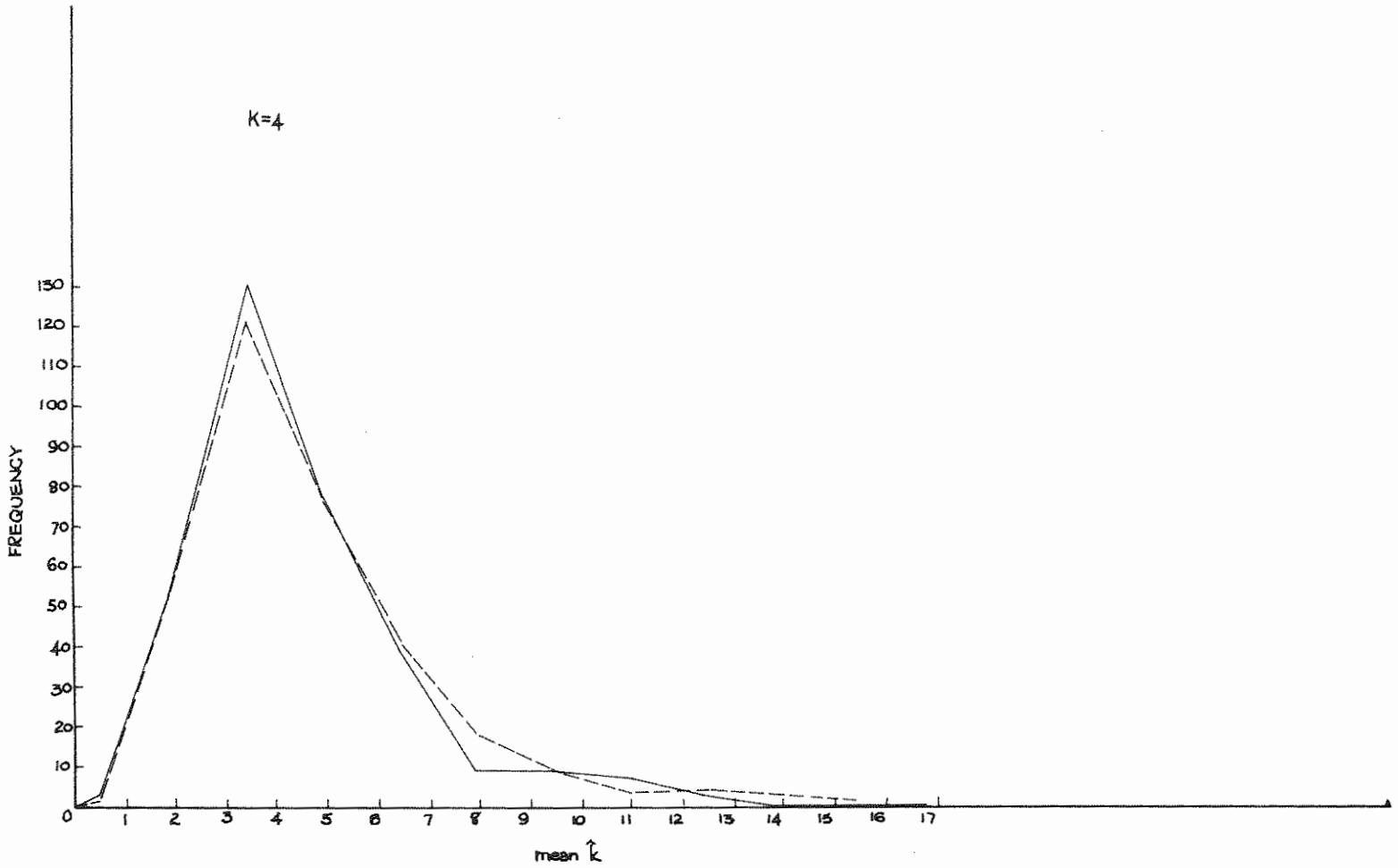

K=1
FREQUENCY
430
420
410
400
390
380
370
360
350
340
330
320
310
300
290
280
270
260
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0 4 8 12 16 20 24 28 32 36 40
mean $\hat{k}$

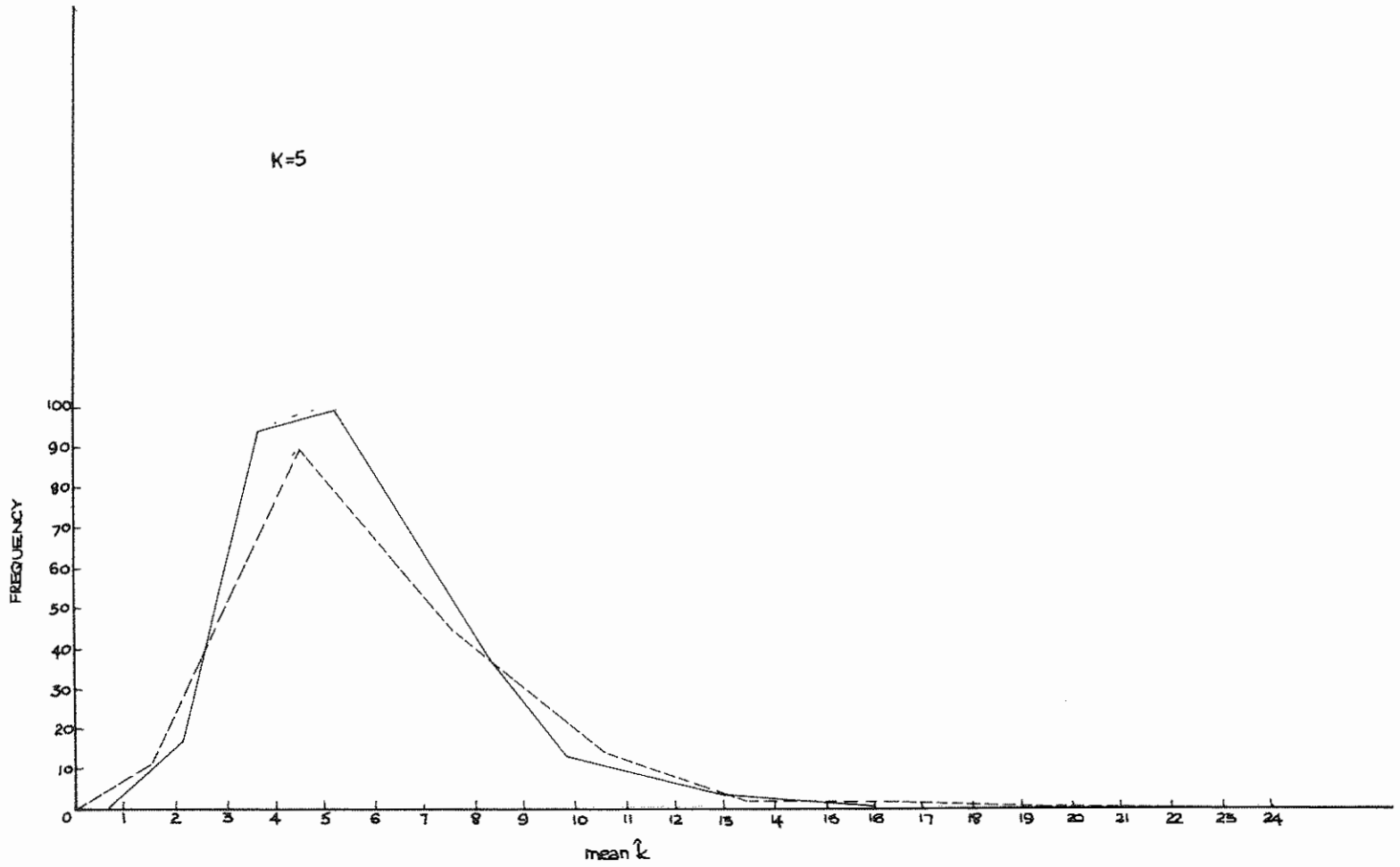
K=2

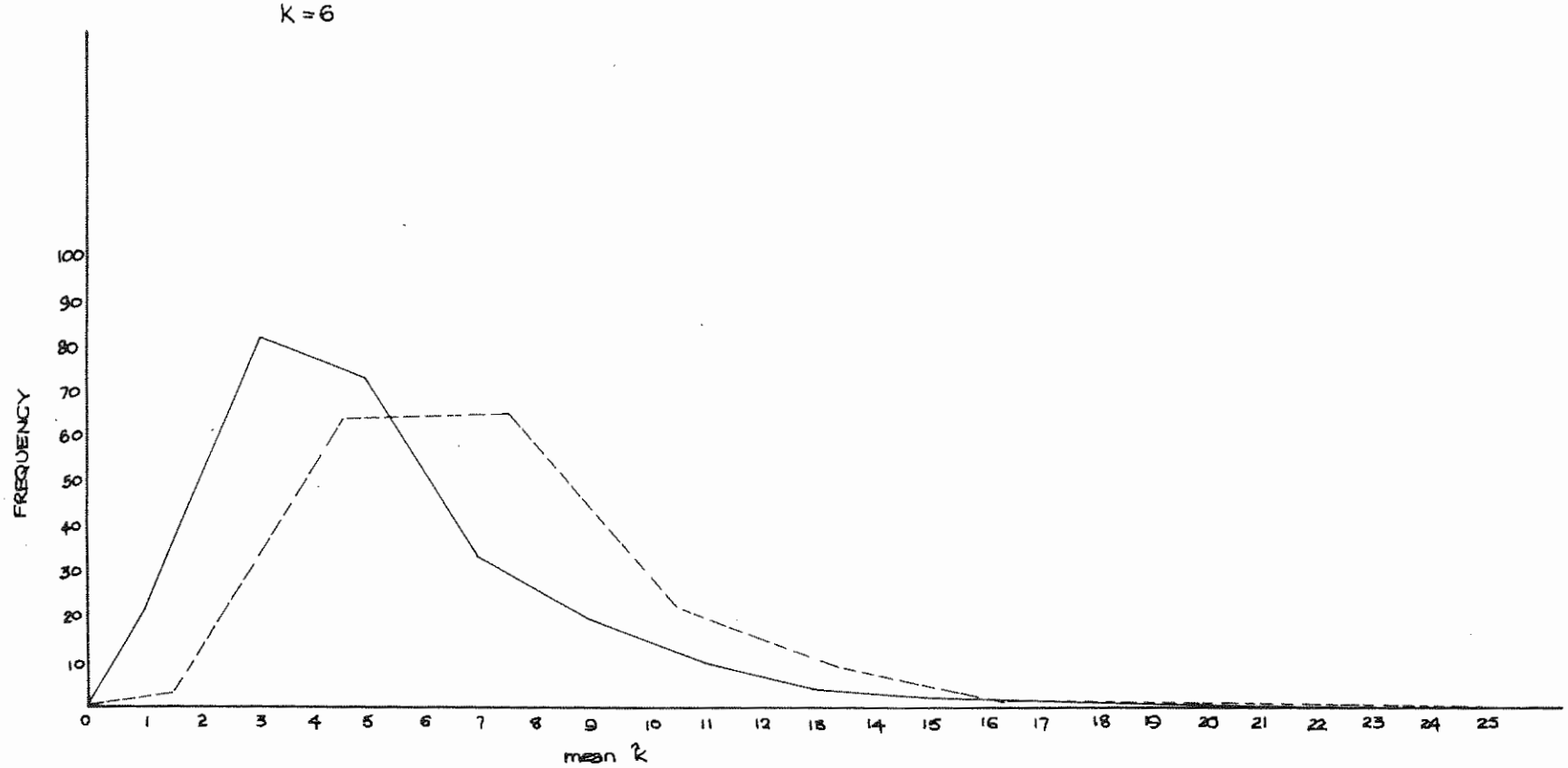

FREQUENCY
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
1·0
2·0
3·0
4·0
5·0
6·0
7·0
8·0
mean $\hat{k}$

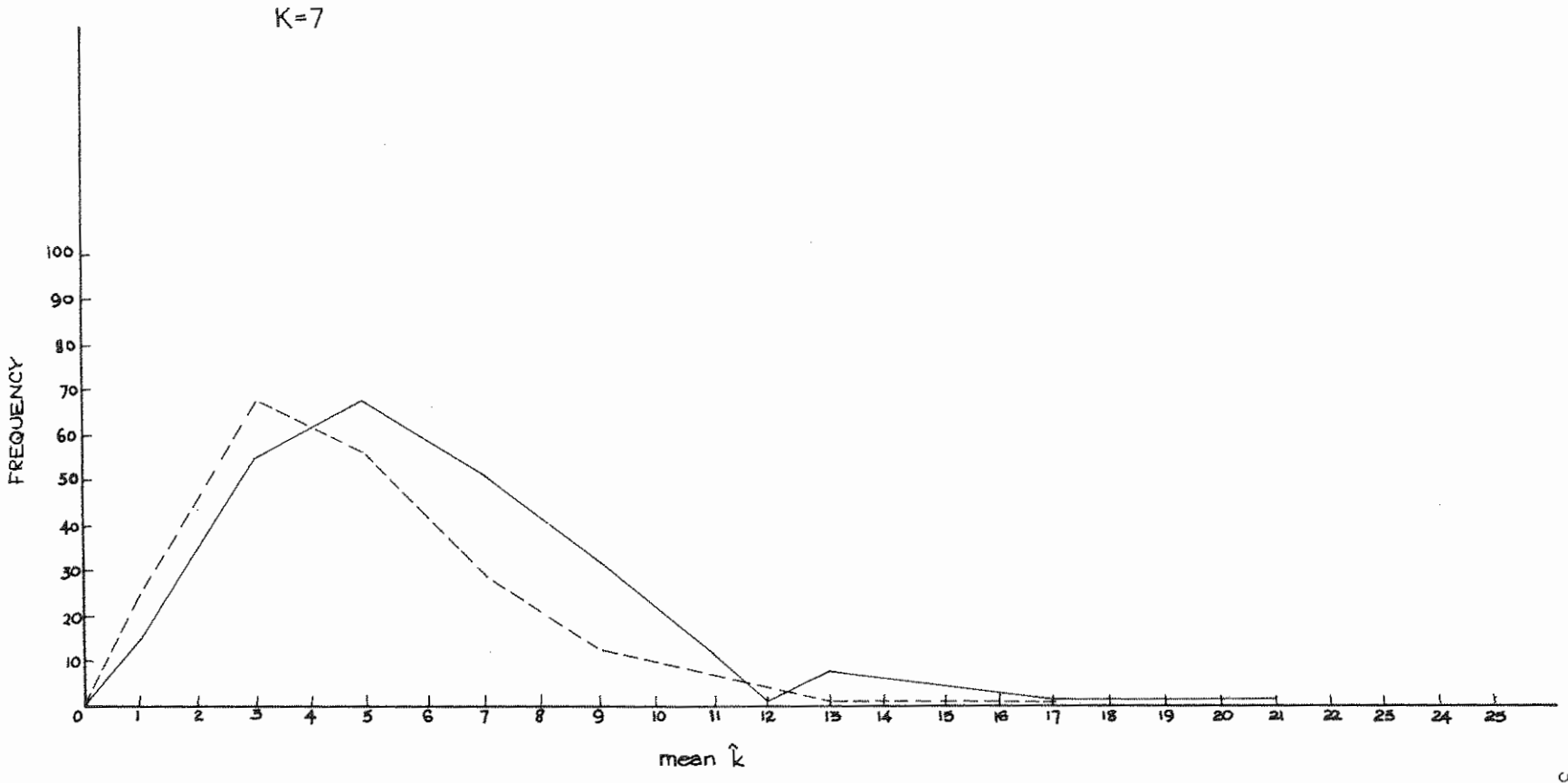
K=3


FREQUENCY
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
mean $\hat{k}$

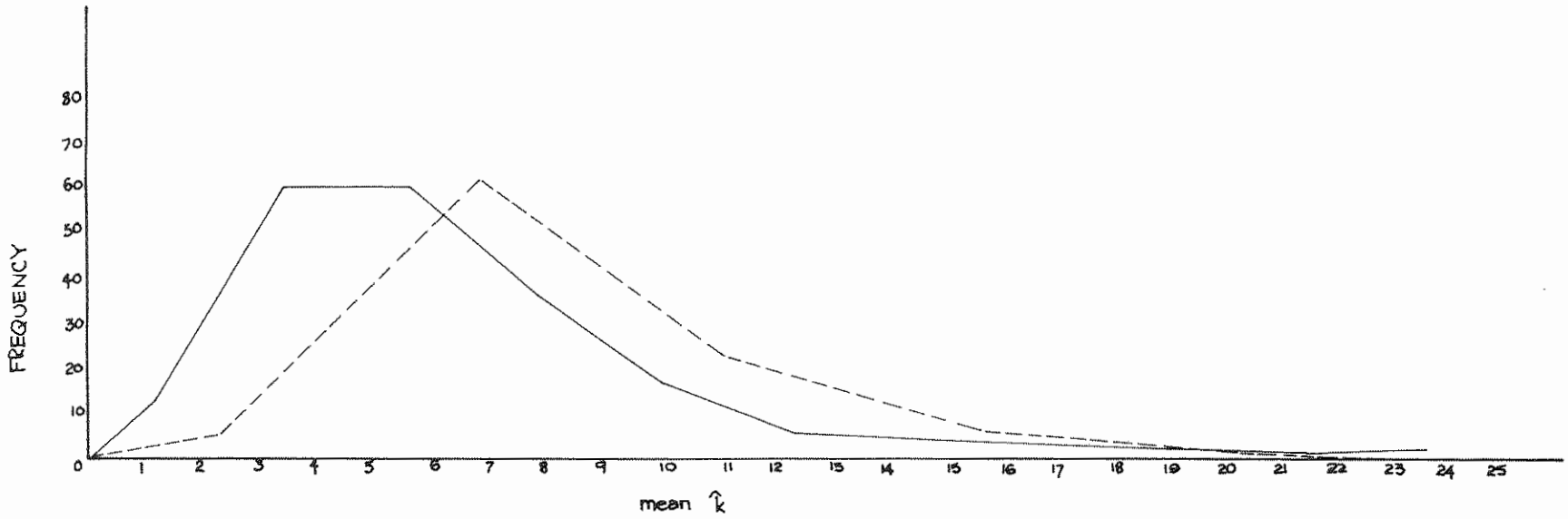

K=4
FREQUENCY
0
10
20
30
40
50
60
70
80
90
100
110
120
130
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
mean k̂


K=5
FREQUENCY
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 16 16 17 18 19 20 21 22 23 24
mean $\hat{k}$

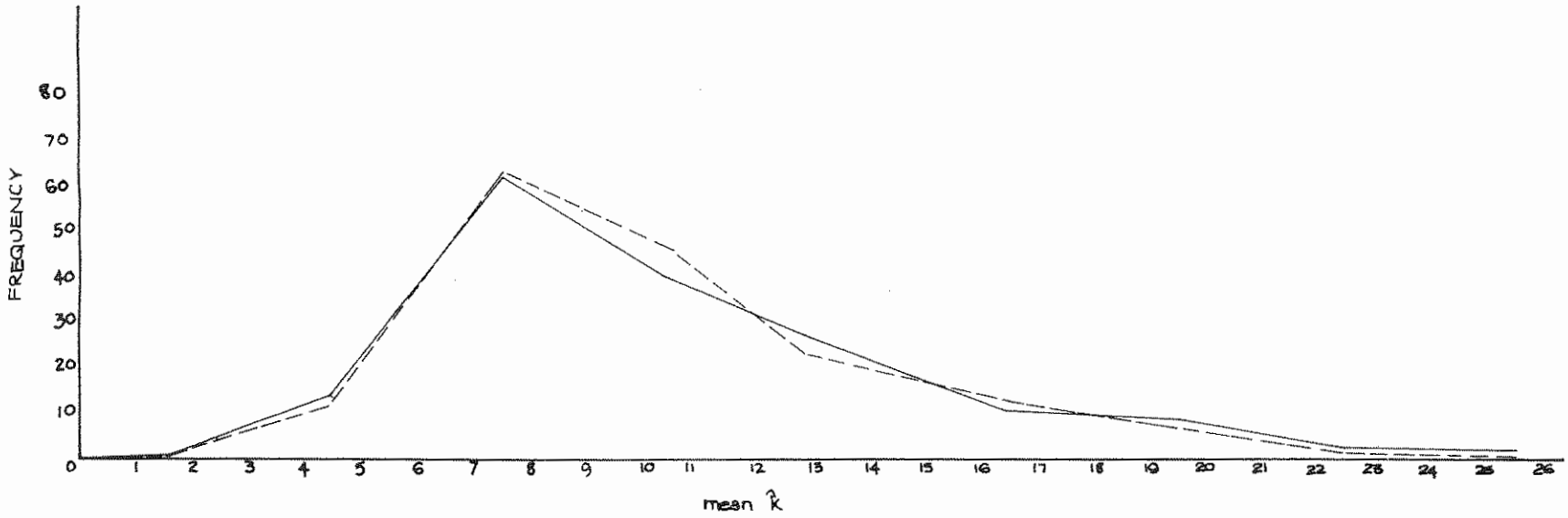

K = 6
FREQUENCY
100
90
80
70
60
50
40
30
20
10
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
mean k

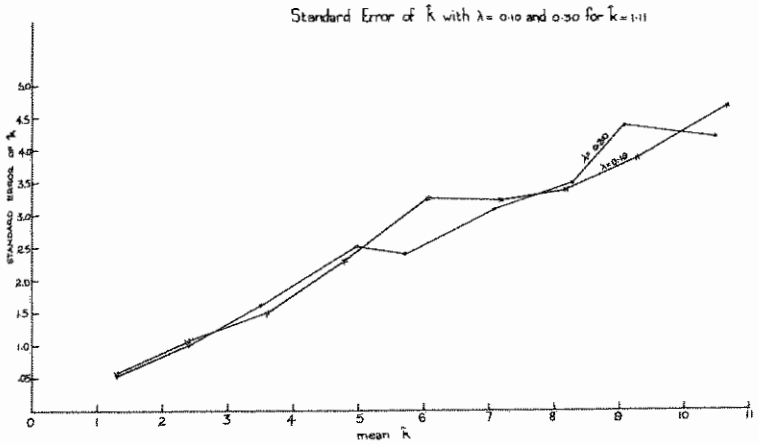

K=7
FREQUENCY
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
mean $\hat{k}$

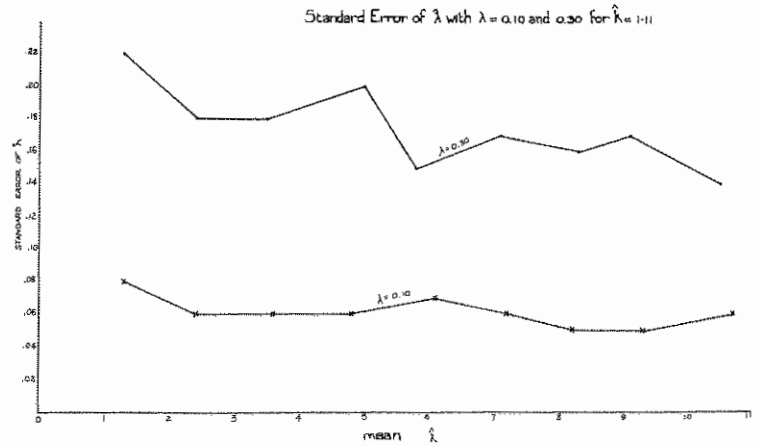
K=8


FREQUENCY
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 15 14 15 16 17 18 19 20 21 22 23 24 25
mean $\hat{k}$

K=9


FREQUENCY
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
mean $\hat{k}$


Standard Error of $\hat{k}$ with $\lambda = 0.10$ and 0.30 for $\hat{k} = 1\text{-}11$
STANDARD ERROR OF $\hat{k}$
mean $\hat{k}$
$\lambda = 0.30$
$\lambda = 0.10$

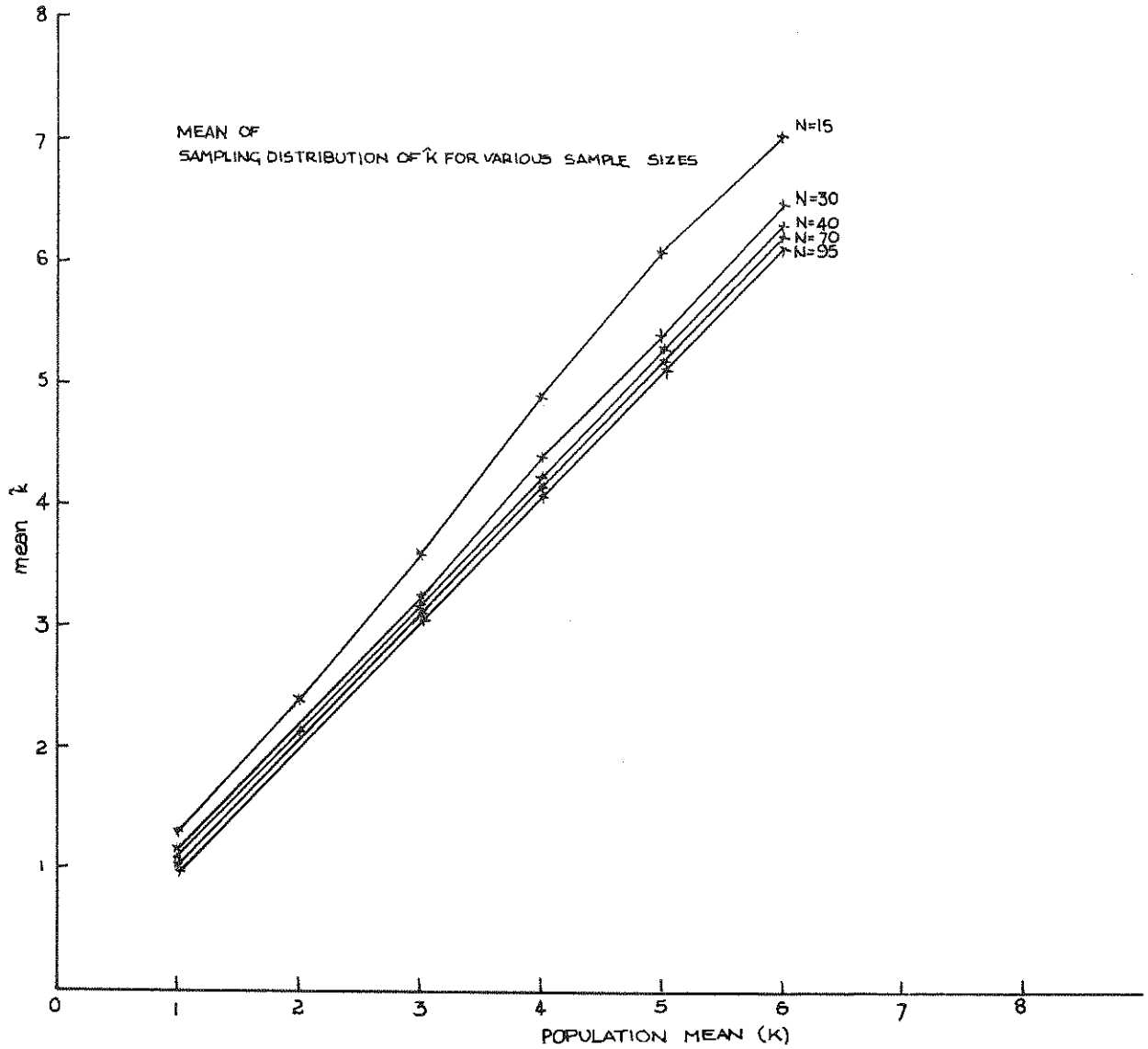

Standard Error of $\hat{\lambda}$ with $\lambda = 0.10$ and 0.30 for $\hat{k} = 1\text{-}11$
STANDARD ERROR OF $\hat{\lambda}$
mean $\hat{\lambda}$
$\lambda = 0.30$
$\lambda = 0.10$

APPENDIX E.

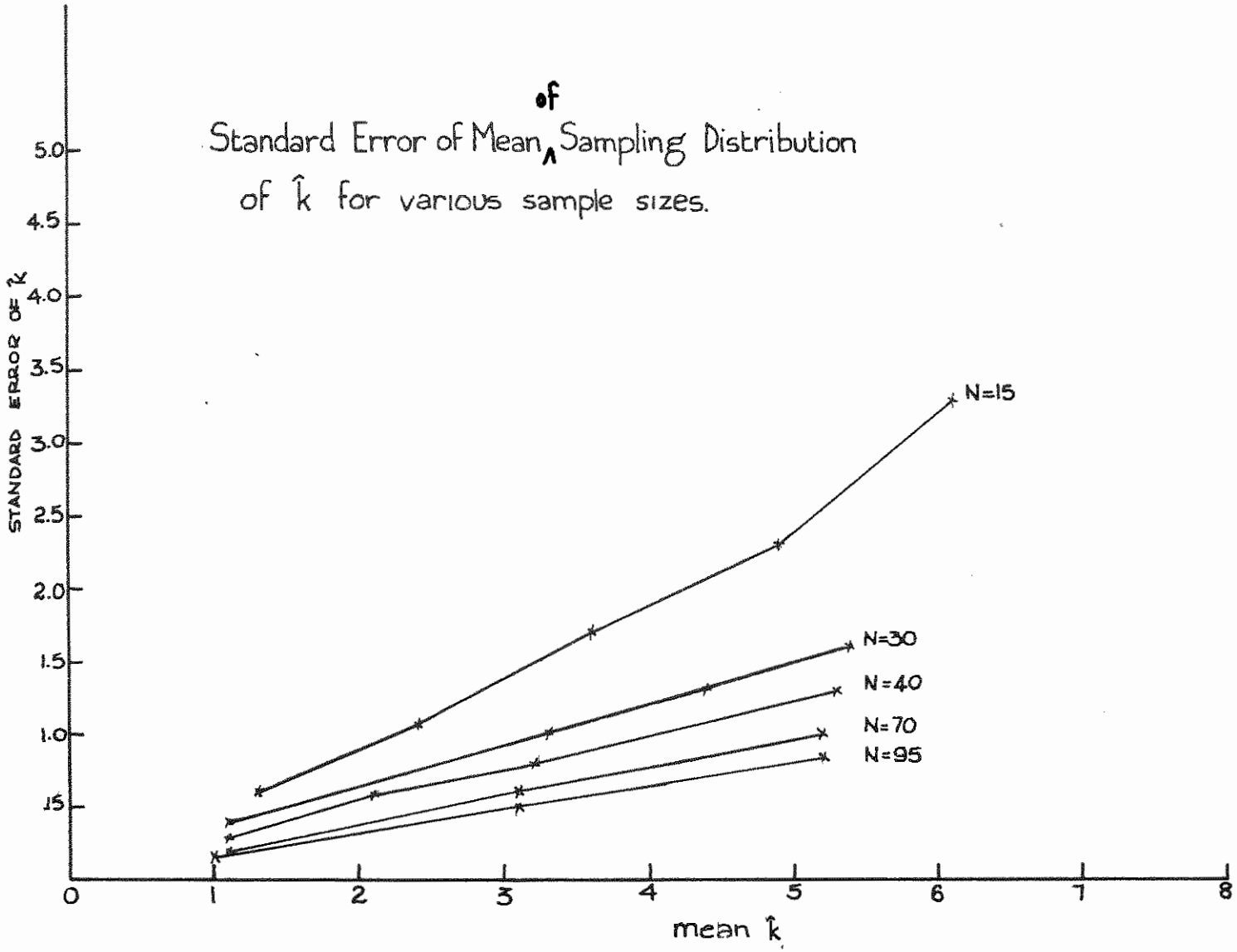
Graph showing mean of sampling distribution of $\hat{k}$

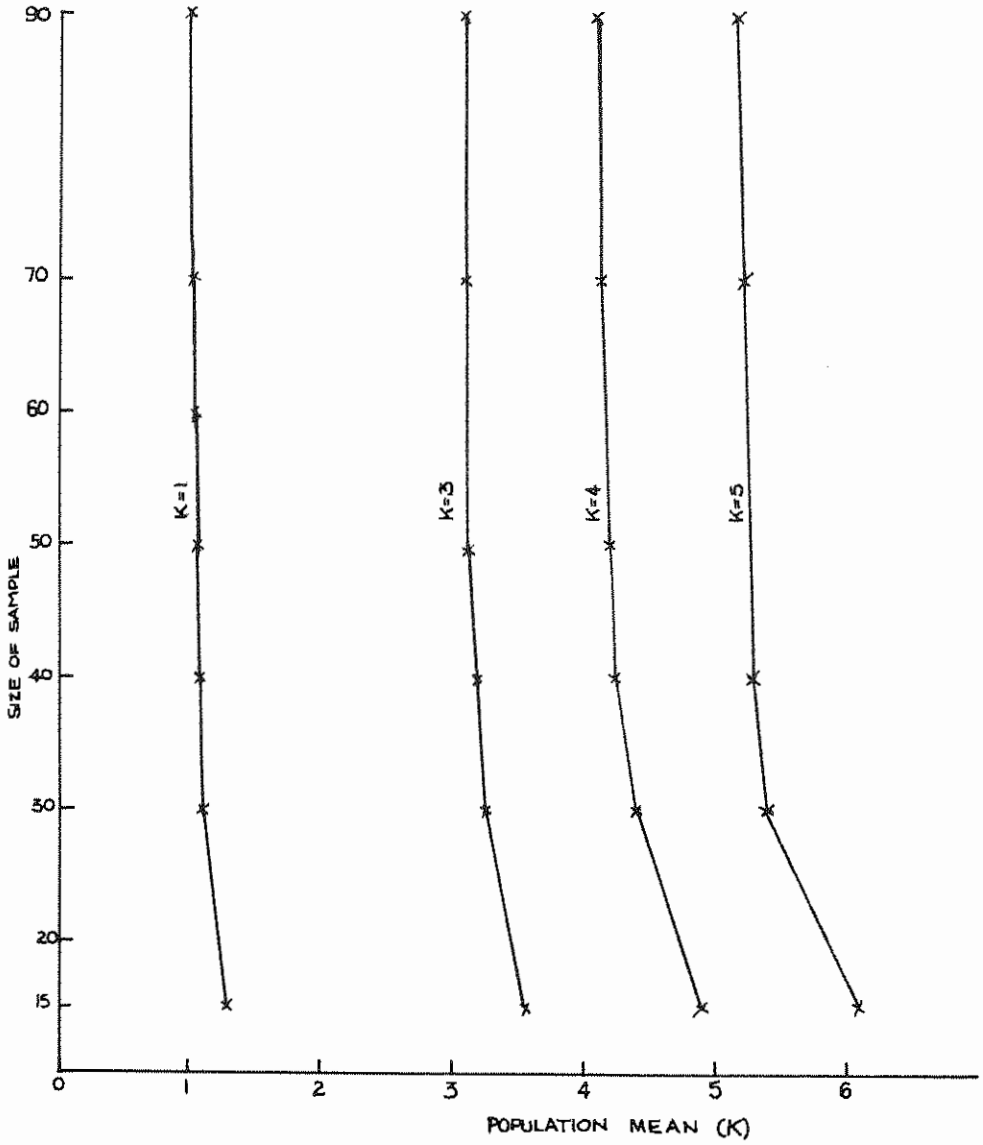
Graph showing standard error of mean of sampling distribution of $\hat{k}$

Graph showing degree of bias in mean of sampling distribution of $\hat{k}$
each for N = 15, 30, 40, 70 and 95


MEAN OF
SAMPLING DISTRIBUTION OF $\hat{K}$ FOR VARIOUS SAMPLE SIZES
N=15
N=30
N=40
N=70
N=95
mean $\hat{k}$
POPULATION MEAN (K)
0
1
2
3
4
5
6
7
8

Standard Error of Mean of Sampling Distribution of $\hat{k}$ for various sample sizes.





DEGREE OF BIAS IN
MEAN OF SAMPLING DISTRIBUTION OF $\hat{K}$
FOR VARIOUS SAMPLE SIZES

APPENDIX F

TABLE I

Mean and standard error (S.E.) of sampling distribution of $k_1 - k_2$ for several sample sizes (N)

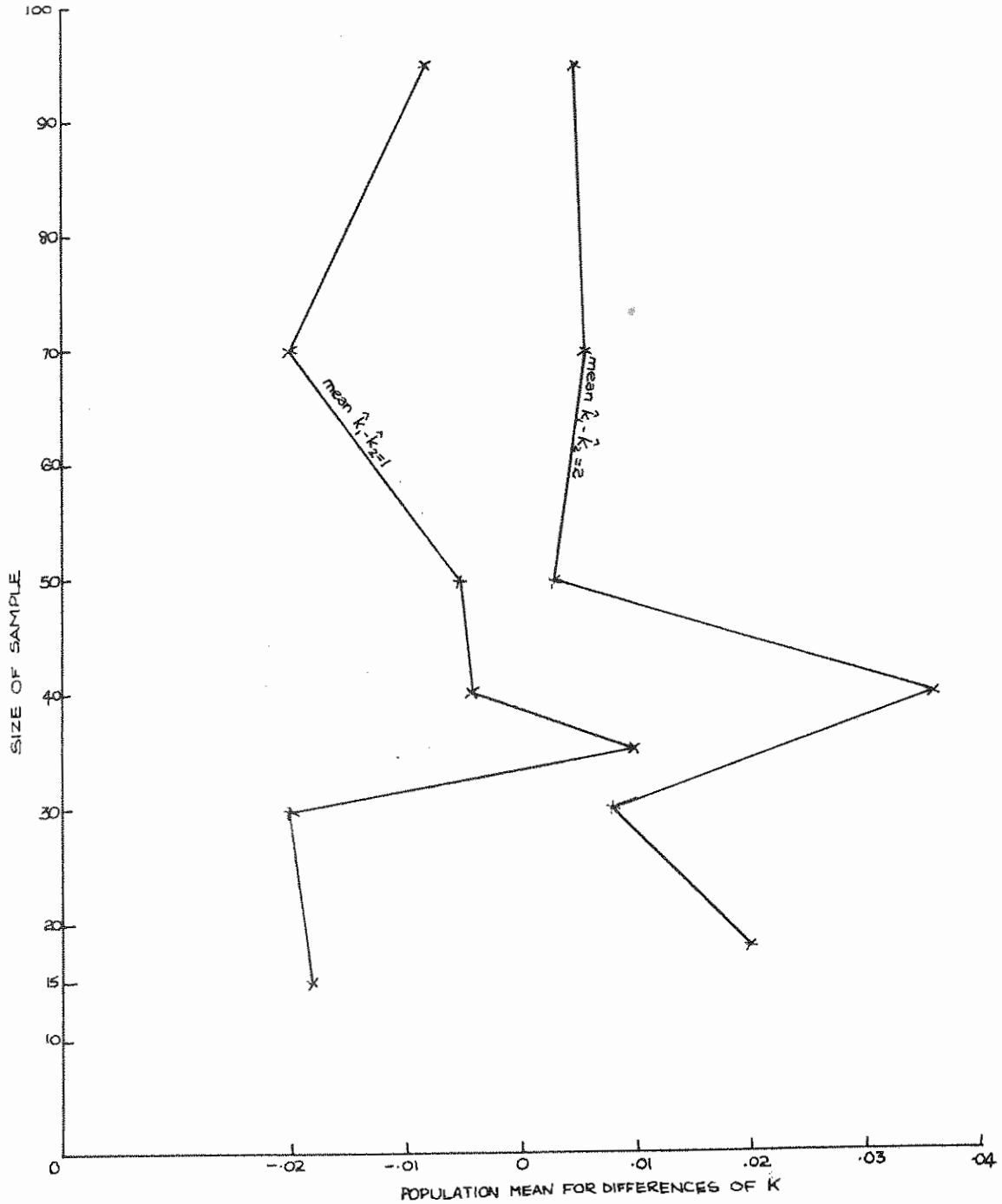
K_1-K_2 =	N	Mean $\hat{k}$	S.E.
1	15	-0.0179	0.7930
	18	0.0139	0.6960
	30	-0.0216	0.5291
	35	0.0110	0.4730
	40	-0.0040	0.4368
	50	0.0055	0.4019
	70	-0.0164	0.3391
	95	-0.0085	0.2834
2	18	0.0225	1.3888
	30	0.0086	0.9710
	40	0.0357	0.8198
	50	0.0031	0.7108
	70	0.0066	0.5908
	95	0.0055	0.5103
3	15	0.1069	2.2056
	18	0.0005	1.9066
	30	-0.1125	1.4184
	40	-0.0239	1.1830
	50	-0.0539	1.0364
	70	-0.0162	0.8367
4	15	0.0027	3.1011
5	15	0.0853	3.7614
6	15	0.1838	4.7568
7	15	0.2246	5.6370

TABLE II

95% confidence intervals for differences of $\hat{K}$ and $\hat{\lambda}$

K_1-K_2 =	N	95% confidence interval ($\hat{K}$)		95% confidence interval ($\hat{\lambda}$)	
1	15	-1.7333	1.6000	0.0260	0.3333
1	60	-1.2000	1.1100	0.0470	0.1949
1	92	-0.6100	0.6400	0.0600	0.1665
3	15	-5.1899	4.3890	0.0399	0.2790
3	32	-3.3000	2.4500	0.0525	0.2050

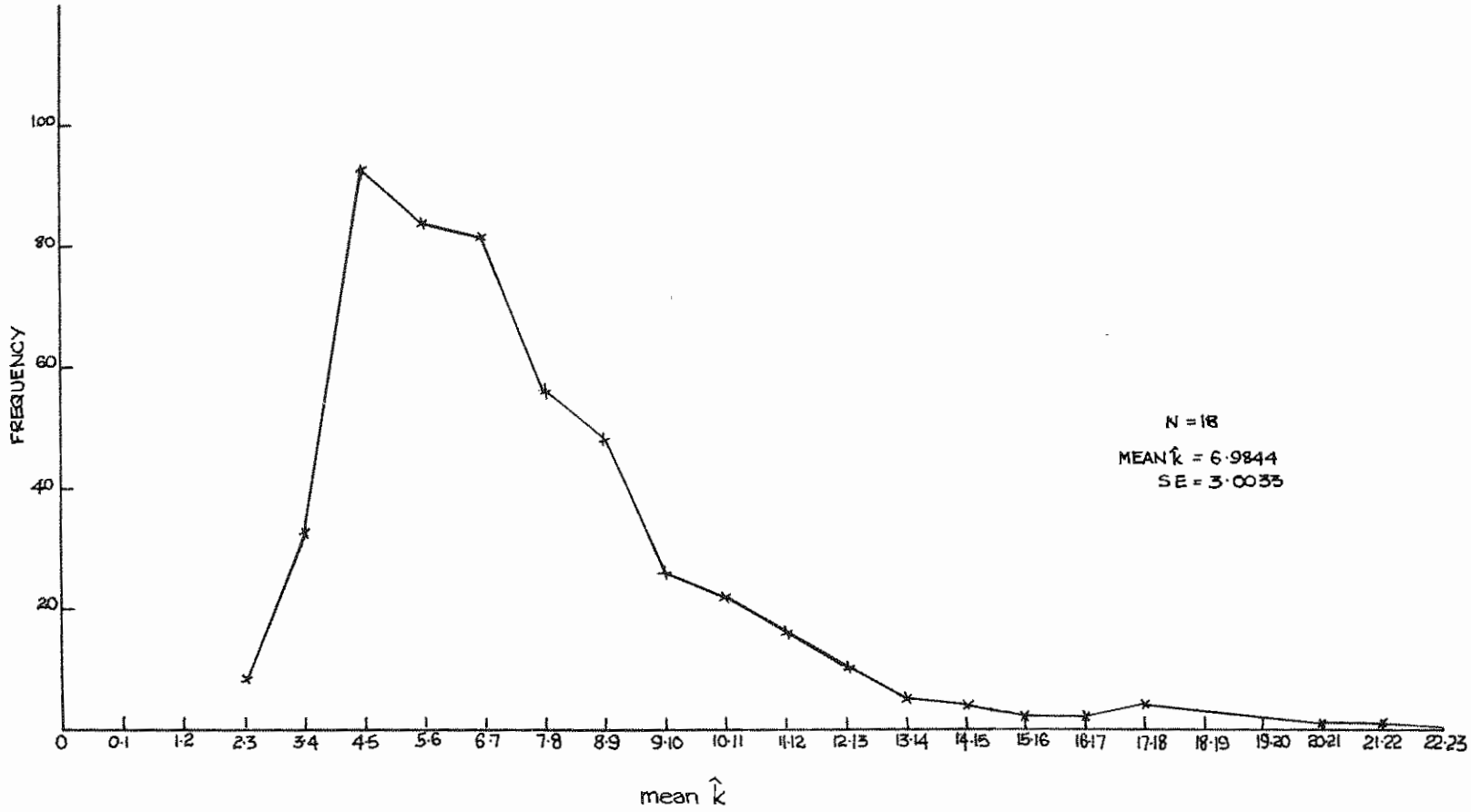
MEAN OF
SAMPLING DISTRIBUTION OF
DIFFERENCES OF $\hat{k}$ FOR
VARIOUS SAMPLE SIZES

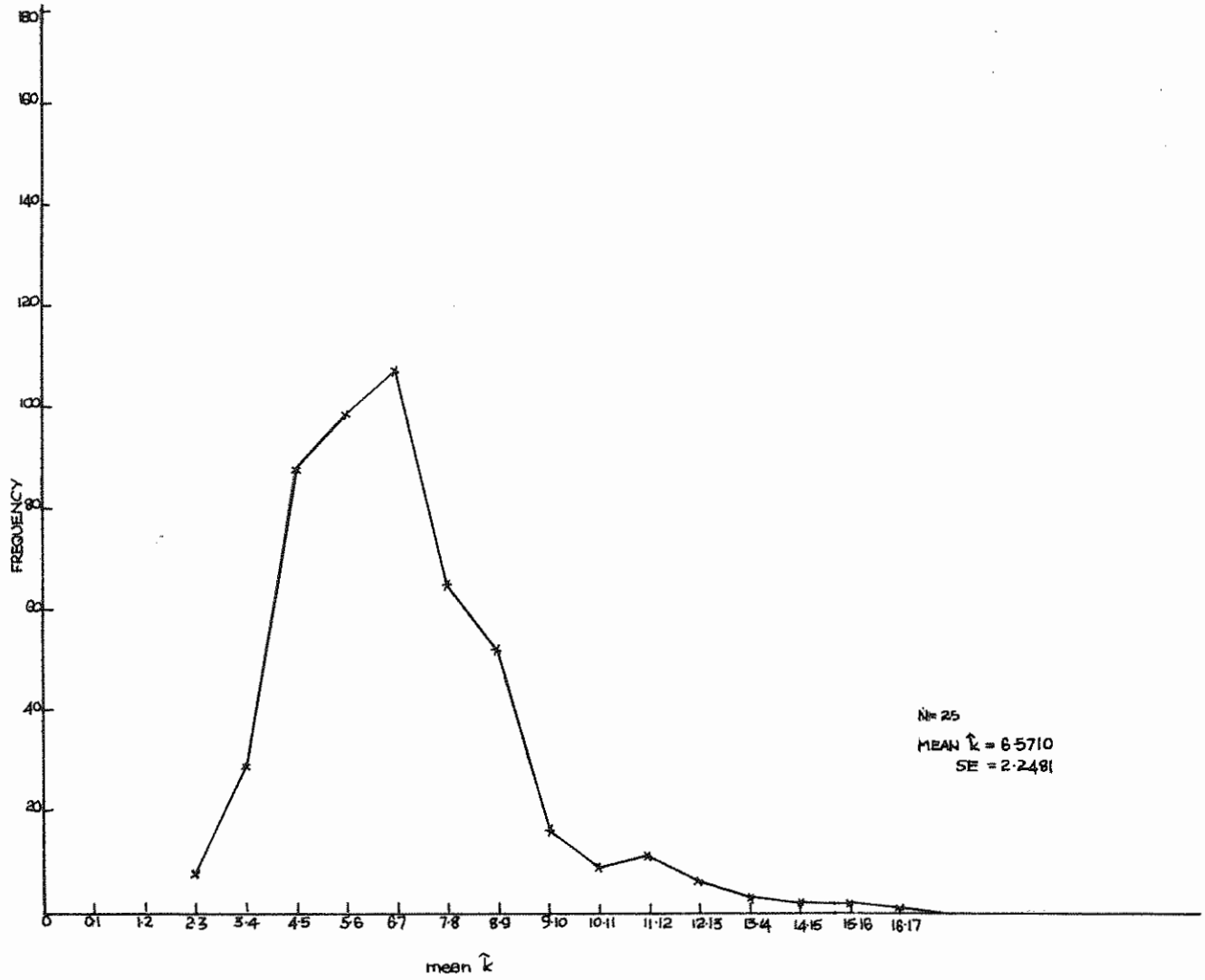


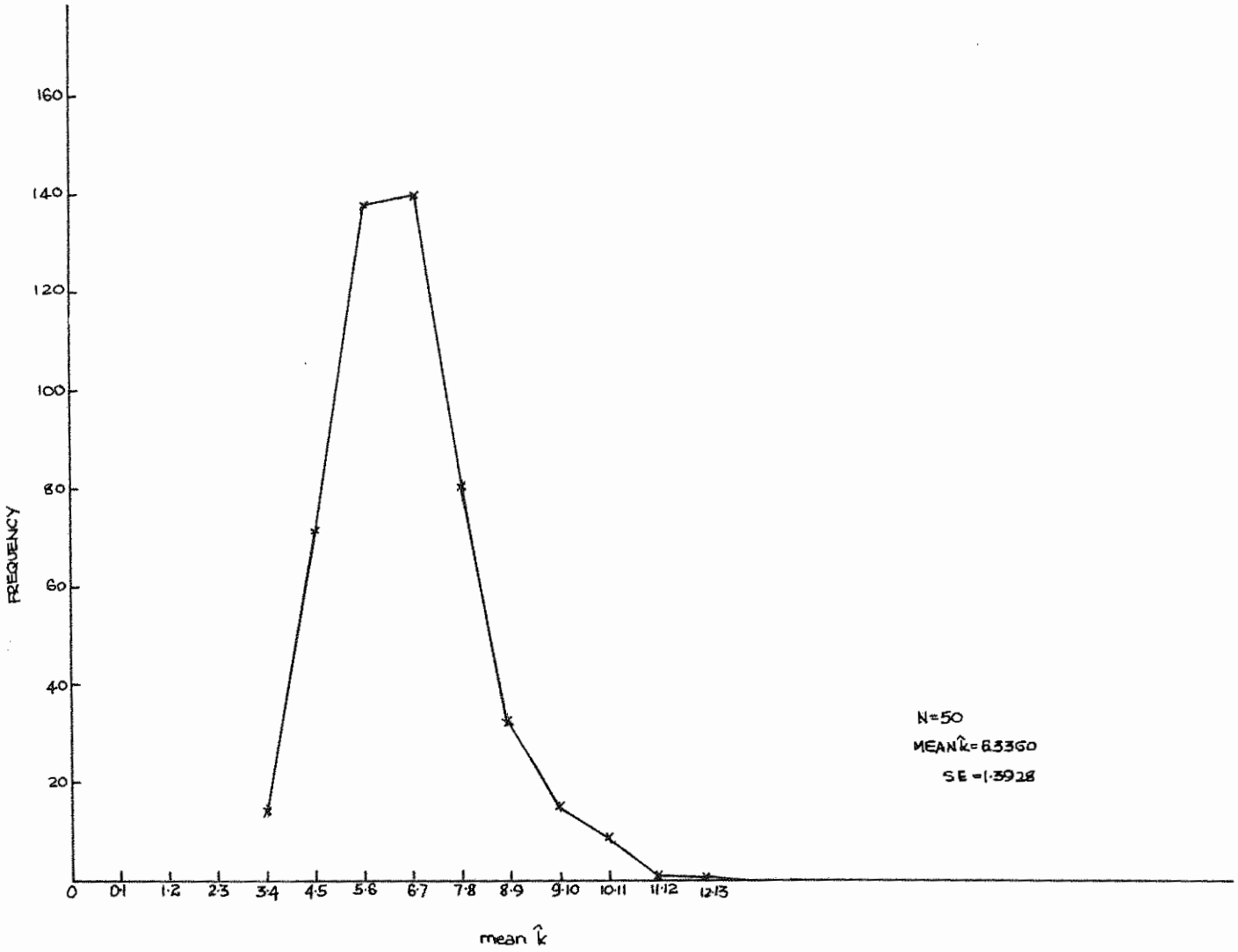
APPENDIX G

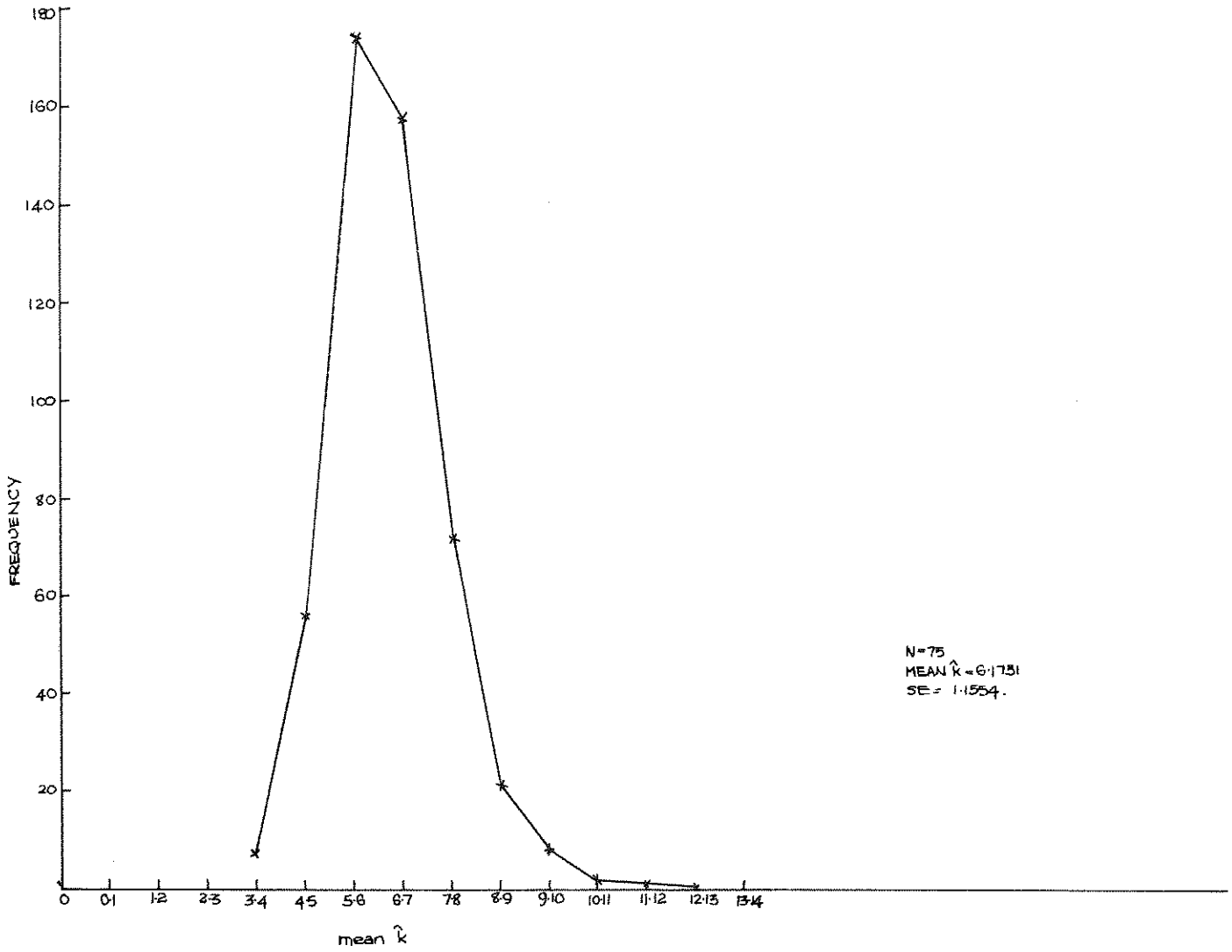
Graph showing sampling distribution of $\hat{k}$ when K = 6, λ = 0.10 for N = 18, 25, 50, 75 and 100

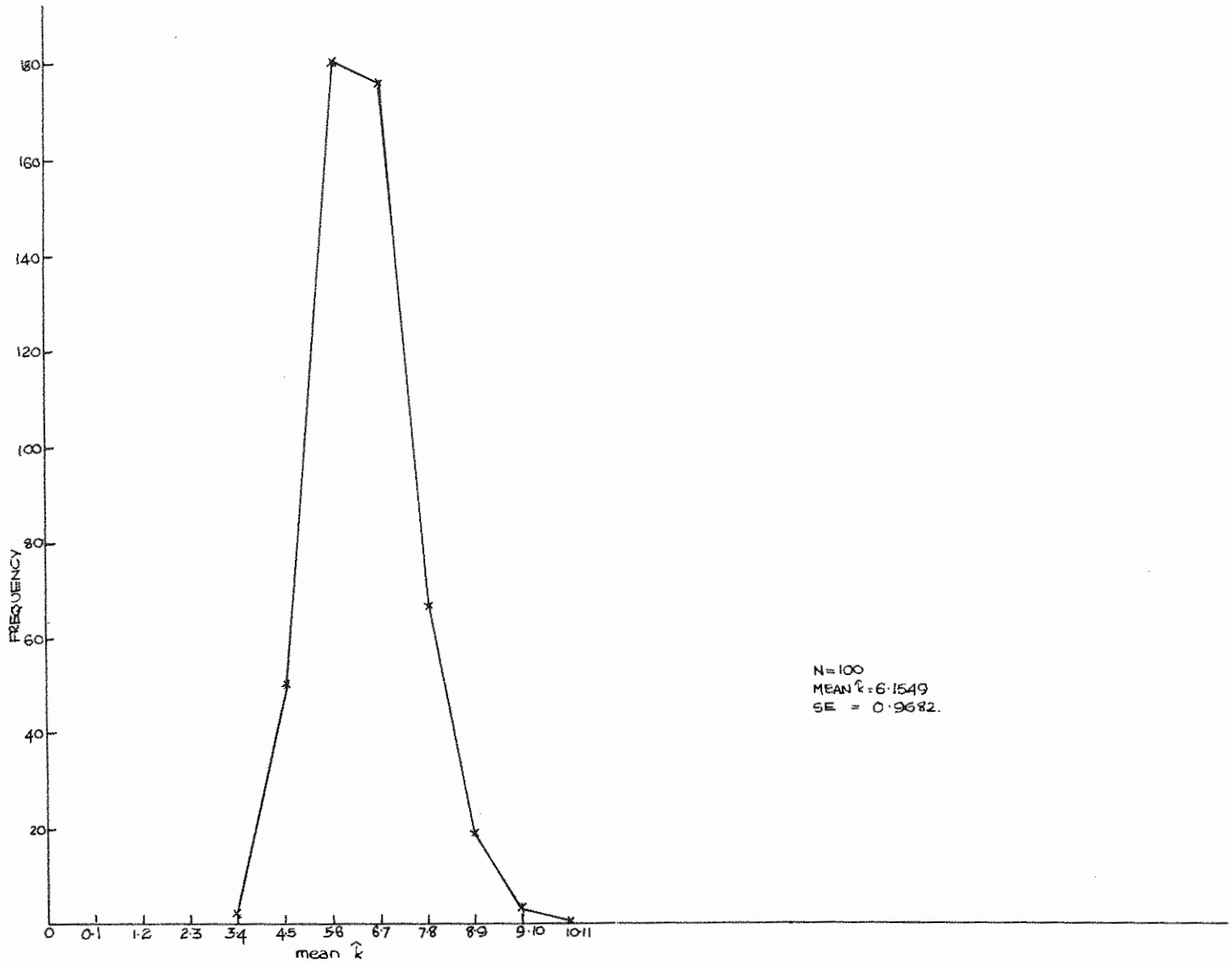
Graph showing cumulative sampling distribution of $\hat{k}$ when K = 6 for N = 18, 25, 50, 75 and 100

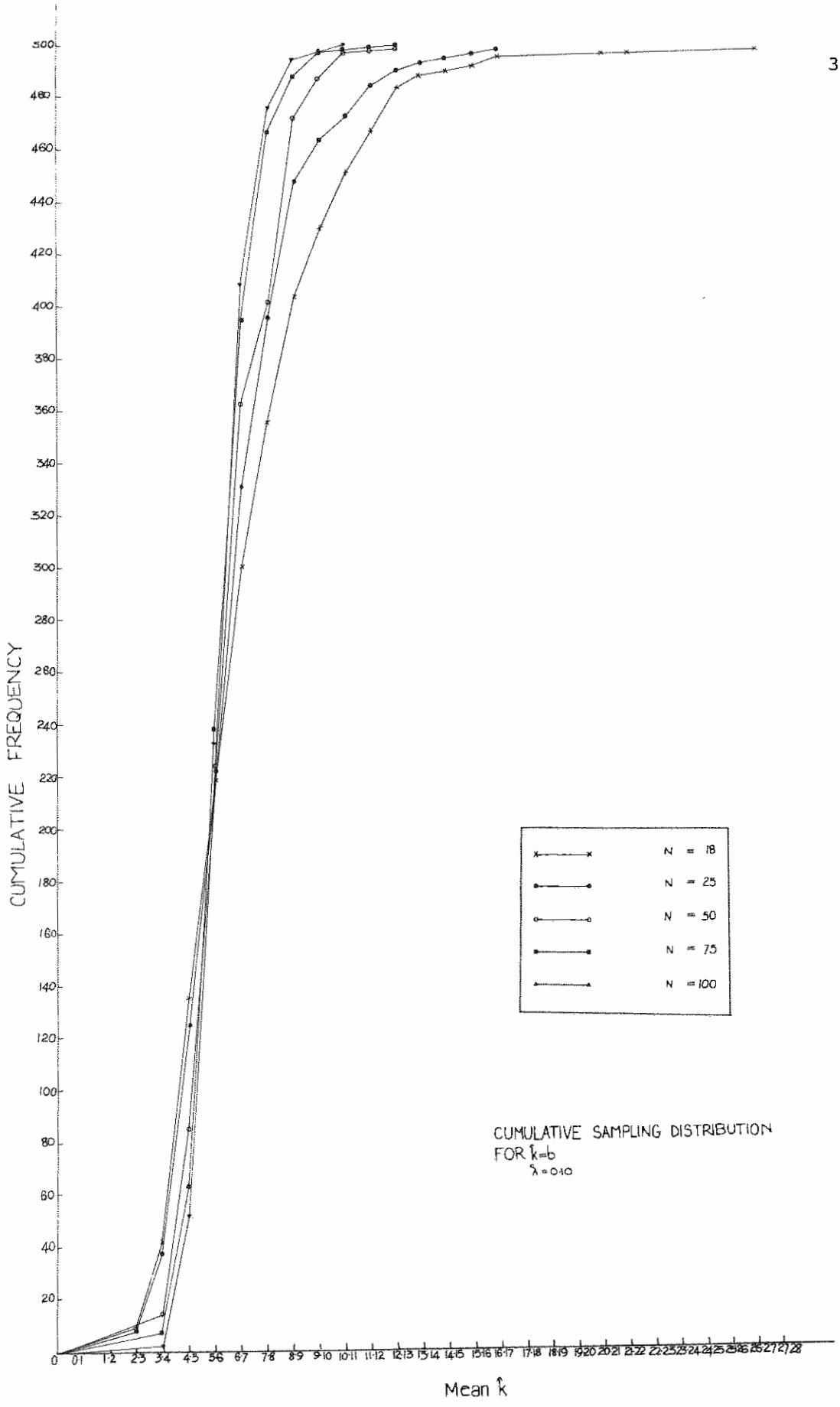

FREQUENCY
100
80
60
40
20
0
0-1
1-2
2-3
3-4
4-5
5-6
6-7
7-8
8-9
9-10
10-11
11-12
12-13
13-14
14-15
15-16
16-17
17-18
18-19
19-20
20-21
21-22
22-23
mean $\hat{k}$
N = 18
MEAN $\hat{k}$ = 6·9844
SE = 3·0035


FREQUENCY
180
160
140
120
100
80
60
40
20
0
0·1
1·2
2·3
3·4
4·5
5·6
6·7
7·8
8·9
9·10
10·11
11·12
12·13
13·14
14·15
15·16
16·17
mean $\hat{k}$
N= 25
MEAN $\hat{k}$ = 6·5710
SE = 2·2481


160
140
120
100
80
60
40
20
FREQUENCY
0
0·1
1·2
2·3
3·4
4·5
5·6
6·7
7·8
8·9
9·10
10·11
11·12
12·13
mean $\hat{k}$
N=50
MEAN $\hat{k}$ = 6.3360
SE = 1·3928


FREQUENCY
180
160
140
120
100
80
60
40
20
0
0-1
1-2
2-3
3-4
4-5
5-6
6-7
7-8
8-9
9-10
10-11
11-12
12-13
13-14
mean $\hat{k}$
N=75
MEAN $\hat{k}$ = 6·1731
SE = 1·1554.


FREQUENCY
180
160
140
120
100
80
60
40
20
0
0·1
1·2
2·3
3·4
4·5
5·6
6·7
7·8
8·9
9·10
10·11
mean $\hat{k}$
N=100
MEAN $\hat{k}$=6·1549
SE = 0·9682.



CUMULATIVE SAMPLING DISTRIBUTION
FOR k̂=b
λ = 0·10

APPENDIX H

Table showing combinations of K, λ range and conditions for Monte Carlo sumulations for which cumulative relative frequency distributions (Monte Carlo) were satisfactorily fitted by appropriate gamma distribution

K	Range of λ values	Condition under which λ was varied
1	1	VaSsSt
1	1	VaSs
1	2	VaSt
1	Constant	CoSsSt
3	1	VaSs
3	2	VaSt
3	2	VaSsSt
3	Constant	CoSsSt
7	1	VaSsSt
7	1	VaSs
7	2	VaSt
7	Constant	CoSsSt

APPENDIX I

Selected Protocols (1-9)

PROTOCOLS

Nine protocols are reproduced in their entirety. They have been selected as examples of both the extremes and the more typical behaviour displayed by the subjects in the study. More particularly, the range of protocols is intended to provide clear examples of the following significant features:

(1) Protocols of all three problems are presented, representing both male and female, adult and student subjects.

(2) Protocols that exhibit regularity of verbal output, as well as a high rate of verbalization (Protocols 1, 2 and 5) are contrasted with those exhibiting irregularity of output (Protocols 3 and 8) and/or a low rate of verbalization (Protocols 2, 3 and 9).

(3) Evidence of personal consistency in both style and rate of verbalization on two problems are illustrated by Protocols 2 and 9.

(4) Other protocols exhibit similarity between two subjects on the same problem in terms of level (richness) of reporting, rate of verbalization and solution times (Protocols 6 and 7). For contrast, two subjects with approximately the same solution times, although on different problems, are seen to differ considerably with respect to degree of verbal output (Protocols 2 and 5).

(5) Protocols were selected to represent different levels of performance as measured by solution time, as well as to demonstrate that no clear relationship exists between mean time and rate of verbalization. For example, Protocols 3 and 8 are from subjects with solution times above the mean for their appropriate age and problem group, yet Protocol 8 shows a high rate of verbalization, and Protocol 3 a low rate. Protocols 4, 5, 6 and 7 were well

below the appropriate mean solution times for their age and problem groups, and display high verbalization rates, while Protocol 9, also representing a below average solution time, has a very low verbalization rate. Finally, Protocols 1 and 2 are examples of average performance in solution time, with a high and low rate of verbalization respectively. (Rate and regularity of verbalization were defined in Chapter 7.2).

(6) Protocols were also selected to demonstrate an absence of use of the display routine (Protocol 2), an extreme use (Protocol 3) and a moderate use in the other protocols.

The protocols have been sectioned by allowing a space between modules. When a definite and correct assignment occurred, the statement expressing this fact is underlined. For Problem 2, dotted lines mark the end of a processing segment as defined in Chapters 6 and 7. Slashes (/) indicate periods of sixty seconds.

Notes on individual protocols follow the presentation of the series of protocols (p.375).

PROTOCOL 1 Problem 1 (W.W. Student;M.) 43.58 minutes.

Shall list the letters with the numbers beside them and work through them by a process of elimination, eliminating each number for each letter in turn, so I can find DONALD, already have D, G, E, R, already have A, L, D. Ø we have, B we don't, E,/R, T's the next. Ten letters, therefore everyone in turn will have a particular digit. Right, we first put down the given data, D=5. Each letter has 8, 9... Right? We substitute. Right now,/knowing that the 2 D's in the first column are 5, that's 10, obviously it directly follows that T=0.

That eliminates, that means we have 2, 3, 6, 7, 8, 9. We'll list those figures we have left, and then we'll put them back into the sum. 5, 5, 10. / Now are there any obvious letters that are self-evident? Er, we can look at the highest column, D, G+R, knowing that D is 5. Er, and the D and G give 5. Therefore, D and G together / D+G must = 5, or if 1 has been carried on, D+G must = 4. Sorry, I've said that wrong. We know that R can only be as great as 9, but D is 5, so 5+G can give us 9 at the most, so the greatest G can be is 4, and so we can only have 9, 8, 7, 6, 5 for the letter R./ So eliminate 0, because T=O. Going to the second column from the right, L,L,R, we know R must be odd, because double L is positive. But we have brought 1 on, so R is odd. Therefore we can eliminate 2 and 4 from R leaving us / with 1 or R+3. Means G...Now we can try a little trial and error here, in the column D, G and R. D is 5, G we don't know, gives R, we don't know. R can be, er/ no, oh the data I gave before for the letter R should have been for the letter G, where G, not R, should be 4 at most. But that still leaves us with the data that R is odd. So

R=1, 3, 5, 7 or 9. We know it can't be 5, so 1, 3, 7 or 9. G is 1, 2, 3 or 4. We survey the problem again with the new bit of information and see if there are any other letters we can eliminate. Giving us,/..no. Now we'll work on the G and R, putting tentatively in for the G 1, 2, 3 and 4 and for the R, 1, 3, 7 or 9. We'll add several values to G and see if we can get values for R and if we can't, we'll eliminate them. So we have first, 4+1, 5+1 give 6. So G can't be 1, because 6 isn't one of the figures we've eliminated for R to be. We see that 2+5 is 7, so 2 could be possible. 3+5 is 8, so 3 can't be./ So we eliminate G to be 2 or 4, and by a similar reasoning we can say that G cannot be 1 or 7. So that doesn't seem to help us a great deal at this stage. So we have to look at some other angles of/ approach for this particular problem.

Trying to work on these double letters such as the double L and the double A. We know that the double L's give an ending of 3 or 9. We know that something is being carried on. Yes, that's right. It's a good clue knowing that L+L+1 carried over from the double D gives R./ We write that little addition sum down by itself to give 3 or L+L+1 to give 9. Now if this were the case, in number one, L would be 1, and in the second case, L would be 4. So we write down if R=3, L=1. If R=9, L=4. This also means/ we might have forgotton something here, the L+L+1 could give 13 or 19, so that would give us some more alternatives. So, if L+L+1 gives 13, that gives us L=6. L+L+1 gives, where are we, gives 19. That means L=8. We have two alternatives for R=3, and if R=3 we now find that L=1 or 6. R=9, L=4 or 8. Now/ can't see that that helps us any. Doesn't seem to at this stage. Work through it again. DT gives 0. Double L, we don't know if 1 is carried over or not, er, so A and A, we can't really work on that. N and R,

we can look at, oh, R's the only one we have any clues to so far. 3 or 9. Going back to R, we, oh, we decided that R had to be greater than 5 I think. Yeh, that's right. D is 5. D is 5+G/ = R. So R has to be, knowing G is not equal to 0, the least G can be is 1, so the least R can be is 6. That's 7, 8 or 9. Ah yes, that's right. R is 6, 7, 8 or 9. Agreed? Yes, that's right. And on the reasoning also that one left, we decided/ yes, we already eliminated G to be either 1, 2, 3 or 4. Now if R, double L, = R, we must have been asleep. R=3, or R=9, that's not right. We know R can equal 9, but it can't equal 3. If we take on LL to equal respectively 6, 7, 8 or 9, we can eliminate 6 and 8, because 2L has to give,/ even plus 1 is odd, so LL+1 gives 7, or 9, or 17 or 19. So correcting what we've already done, working for, if R is 9, L will become as we've said before, 4 or 8, and if L is 7, if L is 7, this makes, sorry if R is 7, L/ =3 or 8. So we write that down. If R=9, L=4 or if R=9, L+L+1 gives 9. L is 4 or if it's 19, L is 9, that's right. But we know it can't be 9. L can't be 9 if R is 9/ so we can eliminate that. So we can say definitely if R=9, L does equal 4. Go over that. R is 9. Try some letter that makes in the first column...Nothing at the moment. D5. G/ 1, 2, 3, 4. Can't really work on G because that doesn't really give us any useful information. Perhaps that can, Jesus, that's a hard problem, er, let's see, ah, here's something. We see here that Ø+E gives Ø. Er, now this means that E must be 0./ Yes, E must be 0, straightaway actually. The only number added to itself give another number is zero. Alright, that gives us E to be, oh wait a moment, we already know that T is 0. Hmm, O.K., so, doesn't seem possible. If T is 0, if/ Ø+E gives Ø, or Ø+E+1 gives Ø. Must be some easier way that that for a number and another, N+Ø+another to give itself. Seems strange. Ø+E gives Ø. 1, 9, 1+1/

1+something gives 1, have to be 10, can't be right. 2=, yes, remember different letters stand for different digits. Er, don't see what it can be, because any number, unless/ zero can be represented by two different numbers, I don't remember if we were told that or not. That makes it rather difficult. Ø+E gives Ø. So Ø + some number gives Ø. Having been told that/ each number is unique, must find something added to something gives x, or something unknown, y, gives x. 5, 10, T is 0. T is 0. TO, must be able to get something into this Ø business. If Ø is 1, let's give it a go and see what happens./ If Ø=1, 1+E gives 1 or 11. Yeh, E could be 9. If E was 9, and we'd carried one on, that would make, no, that would make...correct. E can't be 0, yeh, so that looks good. So E, if Ø is 1+E+another one would give 11. That could be getting, let's have a look at another. 2+E gives 2./ No. Because if that's 12, the biggest E could be is 9, would only give us 11+1 added on is 12. That's 3+E, if that gave 13, ah, could be, if E is 9. Actually on this pattern, it looks very much as if E is equal to 9.

So, let's put that down./ E=9. E=9. Right. now, if that's so, we've already said if R=9, R=9, or R=7. Right? Yeh, so far. Therefore R=7. Gives a big clue. Now we don't know this for sure, not positive, so we'll check it. Go through additions, substituting, 5 and 5 gives 0, carry 1, L+L gives 7 or 17. Right. So that eliminates L to be / if we look back somewhere we'll find it, if we can find it on the paper. 17 or 7, yeh, L=3 or 8. That's right. L=3 or 8. Right? Leave that for a minute. Go back to E being, E, Ø, E+Ø gives Ø, knowing that E=9. We've definitely carried on 1, so with the D+G giving R5. Oh, wait on, we know that D+G gives 7, so G is 1 / because there's 1 carried on. Let's put that back in the problems. So we've G, yeh, we

know G=1. So pop that in. We have 5+1 gives R which is 7 because we've carried one on. Now also we know that N, N+R, N+R which is 7, gives B. Well, there's 1 carried on from the Ø, E, Ø bit, R, 30 minutes already./ Don't know if that helps much. We don't know E, or do we know E? Do we know E=9? Yes we do. We know that E=9. So let's work on this. AA giving E. Now we don't know if 1's carried on to the double A's. We do know that double A equals 9, so that A equals, we know, no we don't know that there's one carried on. Yes, there must be one carried on. So that means, yeh, that's right, R therefore, ah-ha, R must therefore be, or rather the LL/ plus the one carried on definitely gives 19. So we can safely say that L is, L is 8. Can we? Yeh, L is 8. So check that. So we get 5, 5, 10, 8, 8, 16+1, 17, Right, that's correct so far. A+A+1 gives 9 or 19. A+A+1 gives 19. Wait on. Yes, yes, we know that for sure. No, because, or do we?/ A+A+1 gives 9 or 19, so that would make A=4 or 9, but it must be 4. Yes, it must be 4, A must be 4. So put that back in the sum and see if it works. 8,8 gives 17, 4,4,1, is 9. N+R gives B. So that's N+7+1 gives B, we know that./ Let's put, let's eliminate what we got. 5 eliminated, 0, 4,3,4,7,9 eliminated.

So we have 1, 2, 3 and 6. Now we can quickly eliminate some of those. N+8=B. N. We know that B is odd, do we not? Therefore N is odd. So from that we've got left, N must equal 1 or 3. Right, let's try that N=1. N=1+7+ another 1. B would become 9./ but it can't be. Try n=3, 3+7+1 gives 11. Yeh. N=3, N=3, but it makes B=1. Try putting some double A for 8, 1, 9. Right. 8 or 4+1 gives 9, carry 1, N3+7 gives 1, adding 1 gives 11. B=1, Ø, go back and eliminate at this stage what we've got. 3 and 1. Leaves 2 and 6. B=2 or 6./ 7,8,9,10,11. So we know, yeh,

Ø+E gives Ø. Do we know what Ø is? No, we don't. We know that E is 9. Yeh. We already know there was one carried on, so that confirms that 7,10,11, Ø+9 gives Ø. Let's put Ø=2 and see what happens. 1,2,3,4,5. By a little trial and error we shall try Ø equal to 2./ One of the things. We got 2+9+1 gives 12, and we know D is 5. One second, we already said G=1. Continue. Er, up to the stage where that equals that..Somewhere I seem to have worked out that G=1, but that doesn't seem to be right. A is 4, 9,7,18,7, 10,11, G can't be 1 in that case. Ø is 2, 9,10,11,12, if Ø is 6 what happens? Ø is 6,16, yes, this would make, this would be right, Ø would be 6. G's 2. Yes, that looks right. So we shall now, write it all down properly, and see if it's right. D5, D5, Ø6, N3, A4, L8, D again 5, G2/ E9, R7, 485 immediately straightaway. Now we add that up, 10, 17, 9, oops, seem to have made a mistake here. These results can't be right, by the looks of it. 8,17, E is 9, we decided for sure,/ so we'll have to back to the NRB stage. N+R=B. Hmm. Oh, we know straight off R, got it back here, N+7=B. We don't know what N can be, do we? N+7 gives B. We know that/ gives more than,it must be at least 11, because 1's carried over. So N+7 gives 11 or more, N is greater than 4 in that case. Yes, N could equal 4. Well, if it does, B=1. 7, sorry, 4+7 gives 1, Ø+E gives Ø. Ø we thought was 6, yes, right, 6+9 gives 16 and 5 and / 1,5, no that B can't be 1. Hell, B can't be 1. Right. Try N to be 6. If N=6, we have 6+7 gives 13, right? Try Ø to be, what've we got left? We got 2. Yeh, Ø=2. If Ø=2/ that's 13,9,10,11,12,5,1, gives 7, right? Try our sum again, 526,485. G is 1, E is 9, what's R? R is 7. We should know that, heavens above. 4,8,5, add it up, 0,17,9,6,7,13,12,5,6, 7. Now, let's have a look. 5 and 5, want to check we

haven't the same for any numbers/ T1, for 2 we only have 1, that's right, Ø, 3, there's only one, B, that's correct, 4, 2A, correct, 5, 2 D's, 3D's correct, for 6, 1 N, correct, for 7 2 R's, correct, for 8 2 L's, correct, for 9 2 E's, correct, for 0, T, correct. I have finished.

PROTOCOL 2 Problem 1 (K.M. Adult;F.) 17.67 minutes

T=0.

R is greater than or equal to 5, and is odd. Therefore,/ L is equal to 7, no, 3 or 4, no, yes, 3 or 4./ G is equal to 2 or 4. E is even./ R is 7 or 9. Therefore B is probably less than 5. 1,4,6,8/ E is equal to 4,6 or 8. E+Ø=Ø./ If T=0, how can E+Ø=Ø?/ If R=7, E is 2, B is 9, A=1, L=3, G=4, E=2,/ R9, E is 2, must be 8, Al. If G is 4./ N+R is greater than 10. Say N is 6, 6, +7, 13, 3, 1, Ø/ Ø=8, E=9. E can;t be 9 because 2A=E. E is even, say E is 8, Ø is 6, 4, say $ is, no, can't have E as 9, E is 6,/ Ø is odd, Ø is 9, Ø=9, E=6, A=3, and I carry 1, that;s 6, G=1, R=8,/ R is odd. E=2. R=7.

G=1. L must be 3, A can't be 3./ Suppose A is 6, 2 6's are 12, carry 1. Say R is 7, 8+N is, say A is 6/ 2 8's are 16. If Ø is 9, there's 8/ 10+, oh, L could be much larger, 6, say L is 7,/2 7's are 14, 2 8's are 16, 2 9's are 18. Say L is 9, R=8, E is, ...Say L is 8, 2 8's are 17+1=17, R=7, L=8, carry 1, say A=4, then E=9./

R is 7, 7+N=, say 3, N=6, B=3, carry 1, that could be 2./ Right. I have finished.

PROTOCOL 3 Problem 1 (S.H. Student;M.) 55.22 minutes

D=D+D. They're all... D=5, T=0./

Therefore, 2L+1=R. Oh. L+L+1=R/ or 10+R. A+A=E or 10+E/ or A, A+1 = E or 10+E. N+R=B/ or 10+B. A+A+1=E. N+R=B or 10+B or N+R+1=B or 10+B. Ø+E=Ø/ or 10+Ø or Ø+E+1. Suppose E is 9, where do we get to? 9+1 is 19 or 19, right, =Ø or 10+Ø. D+G=R or D+, D+1=R/ Now Ø+E cannot equal Ø unless E=0, oh, can't equal 10+0, D+G=R/ L+L+1=R. L+L+1=R. L+L+1=R, therefore, therefore 5+G=R, and 2L+1=R. 2L, 5+G=2L+1, therefore G, 2L,/ 2L-G=4. Don't know where to start. D+G is less than 10, or R is less than 10/ T=0, 1, G is less than 4./ T=0, E=0. No, sorry, 1, Ø+E is greater than 9, greater than 10./ For sure we know D+D=10. D=5, T=0. R=L+L+1./ L+L+1, R=D+G+1. R=D+G+1. Since Ø is not equal to 0/ R=2L+1=D+G+1, therefore 2L=D+G therefore 2L=5+G=R/ G is less than 4. R is less than 10/ L is/ ah, could be anything. R cannot be greater than 8./ Therefore L is 4, L is less than 4. Therefore L+L+1=R./ L is less than 4. T=0. 2L=5+G=R./ G is less than 4. What? Uh-huh. G is less than 4, less than or equal to 3. Right./ 1 from there, we get Ø+E+1, and that equals 0,7,/ 9,1,17, got to be a 1 up there. N+R is greater than 10, greater than 10./ 2A=E or 10+E, therefore N+R+1 is greater than 10./ R is greater than 5, greater than or equal to 5. Alright, so let's substitute. 5,5,/ 0, A, E, what's E? No known value. N, oh, / 2A=E, 14/ Let N=7,8, let N=6,14. E's got to be 8,9. E's got to be 9.

What?, what? 2A= / E=9, for sure. 5ØNAL5
G9RAL5
RØBERO, D=5/ T=0, E=9.

Now where was my train of thought? 9, therefore A can only equal 4. 4,4. 2L=R, 2L+1=R, R=, let L=3 again, 3,3,1,7,

R=7, then N must equal anything./ 8 say. We go 15,5,Ø,8, 5,Ø,N,A, oh, 5Ø8435
G97435
RØB970/
10,7,it's got to go to 9, E's got to be 9, therefore they've got to be higher than that, they've got to be, make that 3,1,6,3,13,9, 3+8's 11, 11, we'll let that equal 3. Oh, blast./ 4,5,6, say, no, it's an odd number, 7, so let them equal 8, 8+1's 17, 9, 3+8's 11, 10, G equals 2, and Ø=8. No, Ø can equal anything, can't it? G=9, bring in 9, no, it can't equal 9/ 5ØN485
G93485, that's 10, right, 17,9,E must equal 9, pretty well inevitable, 5ØNAL5
G9RAL5, 10/ E+9, O.K., even, A=4, 4+4,what? L+L, where is there an L? Nowhere. There's an R, got to be greater than 1, G can't equal, G can equal 1, can't it? L+L, L=4? No./ L=, L=6, let L=8, 8, that'll come to 17, now if L=9, N,R, got to be a double number. N+R, R's got to be equal to, equal to, what, that equals. The maximum we can carry is 1. Let G=2./ That'll go to 8. 2,2,2 and N. Have we got an N anywhere? R has to equal 7. N must equal, right..52N485
297485/
equals 5,5,10,17,9, carry 1,7, let N=9, we can let it equal, its got to be N's got to be greater than, greater than, N must be greater than 3, let N=3, can't equal 3/ 4, can't equal 4, 5,6,6, can that equal 3? B can equal 3, 12,8. G=1,3,/ can't make it 9,7, oh yes, it can be 7, (whisper) N's changeable to anything. 6,6,13,3/ N's got to be greater than 8 then. Can it be 8 then? Can be 8. 528465
93465, 10,13,9,11,uh-huh, 11,12, good/ R's got to come out to be 3, 7,5,28465
793465,
321930
Tick, tick, tick, now there's no more L's. Tick, tick, 3,3,

no more A's, E's 9, E's 9. Ø is 2, Ø's 2,/ G is 7, R is 3. Alright then haven't doubled up anywhere, have we? D is 5, G=7, R=3, E=9, B=1, don't know/ don't know, N=8, A=4, L=6, D=5, T=0, 0, 1,2,3,4,0,1,2,3,4,5,6,7,8,9,/ Blast, G=7. Right, what can we interchange G with? G we can interchange with N, can't we? G interchange with N. Ah, Christmas. That's doomed it./ Want to change it with something small, that's the only trouble. Oh, 2,7,7,2, oh, D=5, D=5, T=0, E=9,and A=4. G is less than or equal to 3, less than or equal to 3. R is greater than 5. 5,5,ØNAL5
G9449/4L5
RØB9RO

Can't work that out in my head. L+L,/ L is greater than 5, L greater than or equal to, no, greater than, greater than or equal to 5. Let L=5, can't equal 5. L is greater than 5. Let L=7, haven't tried that. 7,7,14,15, that will be alright, no, can't be 7. Can't be 6, I seem to remember. 9, can't be 9. Can't be 8. Can it be 8? Yes, L can be 8. 9./ 9,16,9,17,9,17,5,Ø,N,G,9,7, G=1. R cannot equal 7./ Can't equal 8, can't be 8 or 9 or 7. 5, can't be 7,6, can't be 6,6, 5ØN465
G97465, G,9,13,13,3, ah, that's bad before./ G,9,A,L,5,Ø,R,E, I mean 9,4,4, cannot equal 6. L8./ / A=8./

5Ø6485
197485
7Ø3970 R is 7.

526485
197485
723970, Ø is 2, 17,9,13,12,7,5,5,5,5's, 2,2,2,N,6, A,4, L8, G1,/ E9, R7, B3, 1,2,3,4,5,6,7,8,9/

PROTOCOL 4 Problem 2 (D.S. Adult;M) 7.38 minutes

SIR multiplied by I gives P-double E-P. Hmm. Well, for a start, I x R=P, so I is not equal to 1. I is not equal to 1. Umm, and P is not zero, so also I is not equal to 5./ Umm. Well, umm, what then? Well, let's just try something out, I suppose.

If we had I=2, R=3, say P=6/, and 2 2's would be 4, and 2 S's, no can't be 64. Right. Good, so I can't be 2. O.K. If I=3, now, P could be 1 or 2. If I=3, P can either be 1 or 2./ Well, say if I=3, and P=2 we'd have R as 4. 3 3's are 9, and 1 are 10, so we'd have to have...No, that wouldn't work. 3 3's are 9 and 1 is 10, 1. So 3 S's would have to be 19, look like? Yeh. Good. So that can't be. Right? With I=3/, what was the other possibility? P=1, wasn't it? So that would be R=7. 3 x 7 = 21, 3 3's are 9, and 2 is 11. That would mean E=2. Good, that eliminated P=, I=3.

I=4. If I=4, P could be 2, 3 or 4, oh, sorry, 1, 2 or 3. But it can't be 1 or 3. So if I=4, we must have P=3. 4,2. Umm, that means R could be 3/. Could have R=3 or R=8. Well, if R=3, 4 3's are 12. 4 4's are 16, that's 17, so 4 somethings would have to be 26. That won't work. 8. 4 8's are 32. 4 4's are 16, and 3 are 19./ 19 and 1. 4 7's are 28 and 1 are 29. So 4 x 748 Let's check. 4 8's are 32, 4 4's are 16, 19, 4 7's are 28, 29, so 2-4-7-8-9 are P,I,/ S, R, um, E. That gives me one solution. And any other solution must have I greater than or equal to 6. I have finished.

PROTOCOL 5 Problem 2 (M.N. Adult;M.) 17.20 minutes

The problem is SIR times I gives PEEP./ Well, presumably I is not 1, because then the one digit row couldn't give a two digit row on the bottom. Don't know that there is only one solution, so I'll assume, by the look of it. It's pretty difficult to find a particular solution by just working through, so I think the easiest way would be to assign a particular value to I.

Now, as far as I can see, the easiest value would be, would be 7, I think. I don't know why I'm choosing that number. I think partly because it leaves the value of P fairly clear. It can have a low value which makes it easier. It doesn't constrain the problem very much to take the value of I=7./ Umm, well, that gives 7 in the SIR as well, uh, suppose we take the value of R say as, well it can't be 5, because that gives P5. Uh, we want a fairly low value for P, suppose we take, 9 would possibly be the easiest value. R, 9. 9 7's are 63 and carry 6. 7 7's are 49 plus 6 gives 5, so E is/ 5. Carry 1. Now we've got to have a value of 5 for the next E, and a 3 for P, and you've got to carry 1, which is impossible. Sorry, not 1, got to carry 5. Which is impossible. Have to think about this a little more. You're constrained. Possibly a choice of 9 for I would be better. Then we'll try 7 for R, gives 3 again. Going to have much the same problem as 9./ 9 9's are 81, gives a remainder of, 9 7's are 63, remainder 7, so 9 9's are 81, you carry 8, 38. It's impossible again. Let me see, this seems to be...Umm. Try an even number. I can't see a way of reasoning through this. I'll try 8 for I. That gives I in the SIR, so, 8./ No, the difficulty with using an even number is that you've got to get even numbers on the bottom row. That gives you difficulty in trying to find enough

even numbers. What do we have now? I times R gives P, plus a number times 10. Hmm. I times I, plus this number times 10 gives E. Don't get very far in that direction./ Uh, now perhaps the easiest way would be to assign a value to PEEP, then find a number which divides it. Suppose we take a value 6. No, we can't have even numbers in it, because/... If I is 9, SIR must be less than 800, less than 900, since S can't be 9, and if I is less than 9, if S is/ greater than 900, I must be less than 9. In either case the maximum value of P is 7. Since 9 8's are 72, you can't carry 8, that's not quite true. Yes, you can't carry 8, because you'd have to form that in the same way, and eventually you'd get to the right-hand side of the problem, and you've got to find, from the 9 times something gives something, plus 8 times 10. Which is an impossibility. So P can't be greater than 7. Uh, on the other hand, I can't be 1, because then SIR would have the same number of digits as PEEP, or it would be the same as PEEP. In fact, uh,/ if I were 2, P would of necessity be 1, considering the left-hand side of the problem, since 2 times S can't be greater than 1. But P can't be zero, therefore it must be 1. Therefore, R times 2 is 1, R times 2 is 1 plus something which is impossible, since R times 2 is even, 1 times 10 plus something is odd. Therefore, I is equal to 3, 4, 5, 6 or 7. Not nearly so many constraints on the 3 by the look of it./ Well, we'll consider the 3 case then. 3 and 3 in SIR, uh, yeh, that 3 times 3 gives 9 for E. Suppose something is carried. If it's 1, it gives zero, and a 1 carried. If it's 2, it gives 1 and the 1 carried. Also P must be 1 or 2./ Also, since 3 divides it, P+E+E+P must be divisible by 3 by the test that the sum of the digits of the numbers divisible by 3 are also divisible by 3. Therefore, 3 divides 2P and 2E. Therefore 3 divides into P and E. Since

3 doesn't divide 2, 3 divides P and E. Therefore, now, if P is 1, E must be 2, 5 or 8./ P=2, E=1, 4 or 7. We'll take P equal to 1 in the first case. P=1. Uh. E=2, 5 or 8. R x 3 gives 1. That means R must be 7, since no other number times 3 ends in a 1. I think. Yuh, that's right. Or no other digit will add 10 anyway. So that gives 7, 3 3's are 9, that gives 1 as the value for E. Since 2 are carried, makes that impossible./ Suppose then P is 2. Uh, then 3 times something gives 2, must be 4, since 4 is the only number that fulfills the condition. P equals to 1, 4 or 7. Therefore it's impossible for I=3.

We'll try I=4. 4, 4. In this case, 4 must divide PEEP again, so 4 must divide EP. Another condition for that divisibility,/ which means 4 divides 10E plus P. Uh, 10E is even. Therefore P must be even for 4 to divide the total, so P=2, since it can only equal to 0, 1 or 2. P=2, I=4. Therefore, 4 divides 10E plus 2 which means that E must be odd, because 4 divides / 10E+2, means 2 divides 5E+1. Therefore, E must be, E must be odd since 2 divides 5E+1, 5E must be odd. Uh, therefore E=1, 3, 5 7 or 9. Also 4 times R is numbered, it ends in 2. Now the only possibilities are R=3 or R=8. We'll try the 3 case first. 4 3's are 12. 4 4's are 16, carry 1 gives 17, 7 for E / which is a possibility from the earlier conditions. Now we have to have a value: 4 3's are 12, carry 1, 4 4's are 16, add 1, 17, 7 carry 1, so we have to have a number that 4 times it is 26 which is a number that doesn't exist, unfortunately. Then we must try the condition that R=8, which gives 2 also to the value for P, but then we have 4 4's are 16, carry 3 gives 19, again carry 1, got to get a number that's, yes, that 7, so we have the values S=7, i+4, R=8/ P=2 and E=9. 32,4,4,16,1,4 7's, 28,29, yes, I am finished now.

PROTOCOL 6 Problem 3 (A.R. Adult;M.) 11.55 minutes

U=H+1, and, H+1 or H+0. Since letters stand for different digits, it's got to be equal to H+1. Therefore we have T=1, U=H+1./ If T=1, in the first column, H must be equal to 9. Therefore H=9. Therefore we have H=9, U=0, T=1. S+D=9, because S+D+1 must be equal to a multiple of 10, umm, therefore, since U is already equal to zero, T is already equal to 1/, both S and D must be less than 7. Umm, since there must have been a carry, and S and D, S is less than 7, D is less than 7, A has to be big enough to make the carry across, which means that, assuming that the largest carry from the next column is 2, the largest possible carry is 2, umm/, E must be big enough to give the carry, and yet make...E is not equal to 1 or 0. Which means when E=3, S=7, A=6, 5 or 4. A=6, 5 or 4. Umm, umm, it's not, at least, it must be greater than 6,5 or 4. So it must be/ 4,5,6,7,8. 7 or 8. Umm, N+U+U=D.. Now we know there must be a carry of 1 there, so if N is odd, D is even, N odd, D even, D odd./ N even. Just write down what we have so far. We have 9, A,N,N,A, oh, we have U=0. Therefore, D=N+1. Oh, N,O,S,1,S,1,O,D,Y, over 1,O,E,S,D,A,Y. So we have N+1=D./ S+D=9, therefore we have, therefore S+N=8, D+1+H+1, in the 4th row equals S, because there can be no carry off that row. Umm, therefore D+2=S. Add S to both sides, we get 9+2/ +N=2S. So 11+N=2S. So N is an odd number. Uh, let's have a stab at this stage. What happens if N is equal to 3, and D, that makes D equal to 4. It also makes S equal to 5. So we have 5/ A and Y are just assigned values to what is left over, I think. No, A comes into it up here. 3+1+N, 4+M=5. That would make M equal to 1. That is impossible. What happens if N=2? Then S/ =6,

we have 2+1, that's 3. That's impossible either, too. What happens if N is equal to, N can't be equal to 4. If it's equal to 5, what have we? 5+6+N=3, which would make M=7. Now that would appear to go. Substitute 7 for M. We have 9+1+Y=Y, with one over. 6+3+1+A=A, that's one over./ D=6, nothing over.

5+7+1=3, and one over, since that equals 3, 1+3+A=E. Now we've got A, Y and E left over. Now that can have the values 2,4 and 8. So 4, + whatever it is./ What happens if A is 8? Then E becomes 2 and Y becomes 4. So we have, 985589, 7031, 31062, 104, no, 2, 102368. I think I've finished.

PROTOCOL 7 Problem 3 (G.O. Adult;M.) 11.62 minutes

Right. Obviously, T being the one over, has got to be equal to 1. Because A and S can't give us more than 10. We can only carry 1. We've got to have H equal to 9, and U equal to 0./

Let's knock all those values in and see what happens. 9ANNA9 MOS1, S1ODY add them up and you get 1OESDAY. We'd expect 10+Y to give us Y. But what we do know is A+S+D+1=A. Problem is we can carry 20's here./ Why did I decide we could only carry 1? Because we're only adding 2, that's right. A+S+10+E. A+S=, ah, we might have to carry 1. Well certainly N=D+1. Let's see how many letters I've got. I've got 0, 1 and 9, so I want 2,3,4,5,6,7,8 to go. That excludes from their cards T=1./ Well, maybe it's equal to 2, who knows. Guess I've got to try experiment. Let's first of all put Y=2. That get us nowhere. Fast. What gives us the most information? First start looking at.. A and S occur three times each, so maybe if we start assigning values to A and S. O.K. We know that S+D+1, S+D+1,/ S and D and 1 can be either equal to 10 or 20. O.K.? S+D obviously could, 9 and 8 is 17, is the best I've got only is 9, 9+8, 17,15, so it must be 10, so S+D has got to be equal to 9. So far so good. S and D is 9. So what are the possibilities? 2 and 7 or 7 and 2, 3 and 6, 4 and 5. So at least that eliminates 8. So A, Y, N or M can be 8./ A, Y, N, or M. But we decided that D is the same, since we have S+D=9, that N+1 has got to be equal to D or 10+D. Must be more symmetry. A, A+S has got to be equal to E. That's the only other place/ A comes in. No, E+10. Now let's write them all down. S+D=10 could give us anything. N+1=D, don't know, because of all those things we can just do ordinary addition. Let's try something like, let's put

S=2/ O.K. then what does that lead us to? S=2 then, that gives us that D is equal to 7. 3,4,5,6,8 are left. So if N+1=D, and D=7, N must equal 6. If N=6 and &+M=S, then S=2. Therefore that's equal to 12. Therefore M's/ got to be equal to 5. Here we go. 0,1,2,3,4 to go, 3,4 and 8 to go. We still haven't allocated Y but we can leave that to the end. So we've got A+S+1 = E. We know that S=2, so A+3=E+10. Then we get 8+3=11, so E=1 which is bad news. Again, any symmetry? / A+S, A+S, course there wouldn't be. Y'd be the one that was left in that case, Y's going to be the one that's left. Let's try S=3, then D would be equal to 6. 2,4,5,7,8 are left. With S= to 3, D= to 6, then we get N's got to be equal to, N+1's got to be equal to 6, so N is equal to 5. Therefore 6+M=3, must be 13, therefore M's / got to be equal to 7. Solved.

Got a 5 and a 7 gone now. Then we got A+S+1=E, E+10, haven't allocated A yet. S we know is 3. A+4=E+10. Write down. So how about 8+4? A=8, E=2. Gives us Y=4. Let's check that. 985589/ 7031, 31064, 10, 18, 6, O.K. let's code it. I have finished.

PROTOCOL 8 Problem 3 (A.T. Student;F.) 58.92 minutes

Well, H+T must be 10, so that gets 10+Y equals the same value Y. But T in the TUESDAY in the next column, one column greater than any of the others, so T is probably a small value. Try T=1./ When T=1, H=9 and Y can be any value. Try Y=2 and then you're carrying 1 into the second column. 9,8,hmm. HANNAH's reversible, its halfway there. Yes, it's reversible./ Umm, well leaving Y=2 and T=1. O.K. when we've got Y=2, also try a value for A. Now, A+S+D+1 =A, so S+D+1 must be 10, S+D is 9. So that you can carry the, no, so that you can get the A in TUESDAY. Now you've got T=1 and H=9 and Y-2, so try, umm, fairly large value for S. Ah well, say S=5, and D=4./ Now add those up. Oh, that doesn't help with A. We'll have to try a value with A, say 3, because you haven't got a 3. 3, and we've let S=5. And D=4. Carry 1. We've got D=4, so 2U+N+1=4./ Got a 1, got a 2 and a 3 and a 4 and a 5. So try U=6. When U=6, that must make N=1. Just from adding up, carry 1, we've got N=1, and you've got T=1, so obviously A doesn't, U doesn't equal 6. Because we've let T=1, right up at the top. Alright, try U=7. Therefore N would have to be 9 to make any sense out of it, but I've let H-9, so U doesn't equal 7. And if U=8, you're going to have the same problem. I think. 16, 17, that's going to make N=7. But we have, that's alright, N=7 when U=8. That'll be carrying 2 now. We've got T=1 and N=7. So 7,8,9,10, oh, that would mean M=S because N+T+2 which is 10+M=S so U doesn't equal 8 for these same conditions. But I've let H=9./ S now, I have to try, to go back and try a value of U from 5 down. But I've tried S=5, D=4, A=3, so I'll have to try U=2. Oh, but I've tried Y=2 very early on/ and I can't try U=1 because I've let T=1 so I'm on the wrong track. So I will try T=2, because T is

going to be small. Therefore H=8. Oh, wait a minute. I've only tried H=9, T=1 and Y=2. Now I'll try Y=3./ Still, S+D=9 Umm, and I'll consider S=5 and D=4. Umm, I haven't yet got a value for A. I'll leave that because it's not important./ Now this doesn't, this.can't be solved, because if U=2, then N+4=4. Oh, N could be zero. Right, if N=0, that's still working. You don't carry anything. T=1. You've let S=5, so N must be 4. But I've got D=4. Hang on, where am I?/ I've somehow got onto U=2. But U can't be 2, because you get M and D equal. Oh well, we'll let U-4. But it can't be 1 or 3. Oh, it can't be 4 because D is 4. Can't be 5, because S is 5. We'll have to let U=6. Because 6+6/ is 12. That makes N-2. Carry 1. I've got T=1. 1+4, no, 1+2+M=5. So M=2. But N=2, so U cannot be 6 for these conditions. Oh well,/ have to try U=7. When H=9, T=1, Y=3, S=5 and D=4. Which are probably all wrong. Trying U=7. (whispers...)/ I haven't fixed on A yet. Doesn't make any difference. I think it soon will. 15, 15+N equals something ending in 4, say 24. So N must be 9, but I've got H=9. So U doesn't equal 7. Not with these conditions. Try U=8 which is still possible./ Which makes N=7. So that will add up. But obviously U can't be 8 because in the second column H plus something will be U, oh, TU. And if H=9, and T=1, that means that your remainder from the time before will have been 9, which is hardly possible./ So U isn't as large as I've been trying. So when Y equals, if Y=3, this will mean that Y doesn't equal 3, because there's no value that U can take to make sense. So I'll try when Y=4 for the same values H=9, T=1./ Still S+D=9, but since I've let Y=4, S isn't 5, now D=4. Well, I'll try S=3, D=6, D=6. Confusion. Don't need to worry about A yet/ but you'll be carrying 1 for sure. Now U's going to be a fairly small number, because of this second column. Try U=0. That makes N=5, 5+0+0=6. Right,

you're not carrying anything/ Got S=3. This makes, I think, M=7. Because M+5+N+T which is 1 equals 13. M=7. Yes. Got T=1, and you're carrying 1. You got S=3. 3+A=E./ And carrying a remainder of 1. You must, you have to, because obviously 3+A is greater than 10. What values have we got at this stage? We've got T=1, haven't got a value for 2, got S=3, haven't got a value for 4, you've got N+5, D=6/ M=7, haven't got an 8, got H=9, and we've got U=0. So therefore 3, try A=8. When A=8, 8+3=11, so that makes E=1, but T=1. So A is not equal to 8. But if A=4 or 2 which are the only two values which haven't been taken, this makes/ E=7 or 5 and you're not carrying. So, oh, that must mean that U doesn't equal 0. For these conditions. Pity. / Try, oh, I've tried U=2. U=2. Have I? When Y=3, I've tried it, not when Y=4. Try U=3. Can't be 1, because T=1. That must mean you're carrying 3 in the second column which means A+S+ a remainder perhaps from the time before equals thirty something. So you can't get something like that, so U must be less than 2, and H=9. Well, U isn't 1, because I've let H=9, so U must be 0./

Well, if S+D=9, then try S=2, D=7. So the sum becomes / well, why? We'll have to try U=0 again./ Which makes N=6. Which seems to fit. Right. That means M=5, N=6, T=1+M=12, have to carry 1, 2, S is 2. So 3+A=E. And T is 1 / Consider A, well, say A=8. I think I might have it. This is when T=1, S=2, haven't got a 3, haven't got a, yes, I've got Y=4, M=5, N=6, D=7, A=8, H=9, U=0./ Add it up now. HANNAH, 986689. MUST. 5021. STUDY. 21074. Add it up, 8+2, 10, +8, 18, put down the 8 and carry 1, 6+1 is 7, 5+1 is, 11+1 is 12, 8+2 is 10, +1 is 11, 9+1 is 10. What is TUESDAY? T,U, I haven't got a value for E./ Oh blow. S,D,A,Y. If I could just get something equals 3. E. Well that would work out. No it wouldn't. It would mean E=1/ Why did I say S=2? I just

let that happen. But S doesn't equal 3, because of earlier on Doesn't equal 4. Because I've let Y=4. If it were 5, that would make D4. If it was 6, that'd make D3./ I'll try that. I'm still trying U=0 because I think it is. S is 6./ That makes N=2. You don't carry 1. So N+M=5. Can't be 1+4 because you've got them, can't be 2+3 because you've got them. Can't be 3+2 or 4+1 or 0+5. So S is not equal to 6, with this./ I'll try Y=5. If Y=5, I don't see that it makes any difference, because S+D is still equal to 9, so it can't be 4 and 5. Doesn't make any difference further along except for that little thing. Well, try Y=6, H=9./ T=1 and S+D=9. So S could be 5 and D could be 4. Still try U=0. We've got D=4, that makes N=3. Umm./ That makes M=2. 6+A=E. I'll let A=5 because I want E to be 11. In other words, carry 1, but I've got S=5, so I can't have that/ I've got a 1,2,3,4,5,6. I haven't got a 7 or 8, and I've got a 9 and 0. So obviously, 6+7 is not equal to 18, and/ 6+8 is not equal to 17. So, I don't think.. Alright, I'll still try Y=6, H=9, T=1. And now, S=2, D=7, and U=0./ So N'll be 6. S is 2,6,7, 7 and something equals 12. That means M=5 carry 1, S+A+1=E. And I want that to be 11. So S+A=10. But, I've got 1,2, I haven't got a 3, oh, wait a minute, I've got Y and N both equal to 6. So S is not equal to 2/ and D is not equal to 7. Alright, now, I'll try S=3 and D=6. That makes N=5./ Oh, now I can't have 3 and 6 because I've got Y=6. So that's no go. Try Y=6. S=2,3, 4, and D=5. That makes N=4, but I've got S=4, so I'm on the wrong track there. Don't say this means U doesn't equal 0./ No, I'll try Y=7, and H=9 and T=1 and S+D=9 and U=0. So S=5, and D=4,/ oh well, say S=4, D=5 and U-0. That makes N=4. Oooh. Does this mean U doesn't equal 0? Must mean you're carrying 2 in the H+something=U. Well, we'll go back to trying Y=6, H=9, T=1./ U=1, but T=1, so

U can't be 1, and T can't be 2. Because that would mean you're carrying 3. So A+S+1 would be 30 something which isn't quite possible./ Since H+T=10, let H=8, T=2./ I don't quite see how that could be. No, I'm sure H=9, T=1. I've tried Y=2 and I didn't have a solution. And I've tried Y=3 and Y=4, Y=5,6.7. I'll try Y=8./ And I'm, I'm pretty sure U must be 0. So S+D must be 9. There can't be 5, S can't be 5 or 4 because then you get N=4./ Oh, I'll try S=5, D=4. That means N=3. You don't pay back anything. I've got S=5, so that means M=1. I've got T=1 So S is not equal to 5, can't be 4 because then you get N=4. So we'll try S=3, D=6./ So that makes N=5. So 6+ something=13, which is M, so M=7. M=7. Yes. A+S+1=E. But I've got 3+3. Yes, so/ I want, I've got A+4=11, because I want to have to carry 1 in the next go. So A=7, but I've got M=7 so that's not right. Well, let S=2, D=7 and U=0 still. So N will have to be 6. 7+M=12, so M=5, carry 1. A+3=11. A=8./ But I've left Y=8. But I could say Y=3. Therefore A could be 8. That would make E=1, carry 1, 9, I hope. We've got T=1, S=2, Y=3, oh, I haven't got a 4. M=5, N=6, D=7, A=8/ H=9, U=0. HANNAH 986689, MUST, 5021, STUDY, 21073, add it up, 8+2, 10+8, 18, carry 1, 6+1 is 7, 6+5 is 11+1, 12, carry 1, 8+2 is 10+1 is 11, 9+1, 10. Oh, blow, I've got / TUESDAY becomes TUTSDAY. How is it, oh I want my E to be 1, but I can't have that, so A+S is not equal to 11. A+S will have to be 13 or something. How is it I haven't got a value, a 4?/ I'll try Y4. Won't do any harm. And U0, S+D=9 as ever. So S+D won't be 4+5. Let S=2, D=7./ That makes N=6. And I still want U to be 0. N is 6. T is 1, 6+1 is 7+ something=12. That means M is 5, carry 1. A+S=something, haven't got a 3. Well E could be 3./ Oh, that makes A=0 or 10 or something there, which isn't possible. I haven't got an 8. Oh well, S is not 2, D is

not 7./ Try S=3, D=6./ Therefore N=5. S=3, 5, +1, that's 6+something=13. That makes M=7.

I think I've done this before. Carry 1. S=3, 4+A=13,/ that makes A=9. Oh no, I don't want E to be 3. Can't be 4+7. No, S is 3, and you're carrying 1, so you've got 4+A=E and I want the total of E to be 10 or greater. Now the values not taken, we've got a 1, got T=1, we haven't got a 2, so E could be 2, that would make A=8, which might work. Got T=1, E=2, S=3. Try Y=4. Try it./ H=9, A=8, N=5, M=7, U=0, S=3, T=1, S,T,U,D=6, Y=4 and E=2. 1,2,3,4,5, 6,7,8,9.0, HANNAH, 985589, 7031, 316, no, 310/ 64. Add it up, 9,1.10,14,18,6,13, and what is TUESDAY? 1023,S,D,A. Yippee, I've finished.

PROTOCOL 9 Problem 3 (K.M. Adult;F.) 22.50 minutes

T+H=10. S+D+1=10. T=H=U/ =H+1. T=2H+2+M, query last one. U probably/less than 5, therefore H is less than 5, U greater than five. Say H is 3 and T is 7 / / , T is 1, H=9, U=0. No, it can't be. Must be.

D=N+1,/ therefore D+M=S, and S+D=10, S+D+1=10. S+D=9, and D+M/ =S. D is, 6+3, 3,6,7,2, S is 7, D is 2, M=5, 5, D is 1, 6+M, that's S, S is 7, too big./ 8 and 1, if S is 8, N+8, 2+7, 3+6, 4+5, S+E and D+M=S. 2+5 or 3+3 is out of the question. 1+3, can't have that anyway, it's got to be 2+5/ A+S=E, 6/. Where'd I get D+M=S from, and D=N+1, or D+M+1=S?. 2 and M is 4, D could be 2 or 3, and M2 or 3/ makes S6. 4,2,3,S,D./ Say S is 6 and D is 3. If D is 3, then N is 2. 2+1 is 3. M is 3. I said D is 3./ D is 2 and S is 7, makes N1. If D=3, N=2./ If D=3, N and M must be 4 and 2. N is 2, M must be 4./ S=7. No reason why I can't carry on there. S is 7, D=2./ A is 5, 6+5=11+1=12, E=2, N is 3 and D is 4. // D=3, N=2, S=6. Doesn't make sense./ S is 5 and D is 4 and N is 3./ 3+1=4, 4+N=5. N can't be 1. S is 6 and D is 3. N is 2. Makes M 3. Can't be either. S is 8. S is 3, D is 6, makes N5/. 5,1, 6, makes M=7.

Carry 1, 1+S is 4. A must be 8 and E=2. / 9, Y=1, 2, 3,4. 8, 5, 5, 8, 9, M is 7, U is 0, S is 3, T is 1, S is 3, T is 1, U is 0, D is 6, Y,4./ 1...

NOTES ON DISTINCTIVE FEATURES OF INDIVIDUAL PROTOCOLS

Problem 1 (DONALD)

PROTOCOL 1

The time taken by this subject was approximately average for his age group (Mean time Verbal/Student/ Problem 1=38.99 minutes). He is, however, more verbose than the average subject, displaying much of the peripheral behaviour described in Chapter 7.6. On this, and Problem 3 (not included), he constructs a quite elaborate blackboard on which he operates by means of the method of elimination, with some success. A large part of the difficulty he encountered was through failing to realize, between the beginning of the problem and the end of Module 1, that R was not only odd, but had to be greater than 5. He uses the concept of double letters: 'double A', 'double E', which is interesting only because it does not appear in other protocols. Between Modules 1 and 2 he is hindered by not appreciating the significance of the Ø+E=Ø column, and by his insistence that E must equal zero. Once he realizes that E is equal to 9, he solves Module 3 very quickly, assigning the letters in the order R,G,L and A. He takes unusually long to solve Module 4, and this is mainly because he forgets that he has previously assigned the digit 1 to G. Eventually, after calling the display routine, both for elimination and stock-taking, he stumbles on the right combination for the letters N, B and Ø, and completes the problem.

PROTOCOL 2

The information in this protocol is very sparse, an occurrence all the more unusual since the subject was female: other female subjects (see Protocol 8) tended to

be verbose. The spare quality of this protocol suggests the analogy of an iceberg of which only the tip is visible. Nearly every statement represents a new state of knowledge, with very little bridging information about paths taken, or inferences made. The mean solution time for the Adult/ Verbal group on Problem 1 was 16.34 minutes, so that this represents an average performance in terms of solution time. This protocol may be compared with Protocol 5 in which the subject produced at least five times as much verbal output in the same length of time. The subject of Protocol 2 is also unusual in being the only one to following the solution path from R to G, L, A and E (see Figure 6.5).

PROTOCOL 3

This protocol was selected mainly because of the intense use it makes of the display routine. It is interesting also for its irregular, and usually low rate of output. The subject appears to be predominantly concerned with relationships of inequality (i.e. greater than or less than), and either tends to forget assignments he has previously made or considers them to be reversible. (This contrasts with the rigid behaviour displayed by the subject of Protocol 8). Being flexible about previous assignments has heuristic value only if that assignment was wrong. If it was correct, it produces errors and attenuates problem solving time.

Problem 2 (SIR)

PROTOCOL 4

This subject, who had considerable previous experience with cryptoarithmetic problems, solved this problem in considerably less time than the mean time of 20.62 minutes for the verbal group. The extremely short time of less

than two minutes in the first processing segment may be attributable directly to the fact of familiarity, in other words, experience acts to shorten pre-module search time. The subject sets the value of I almost immediately, first at 2, which he abandons readily, then at 3, with P either 1 or 2. He first tests P=2, and when this proves contradictory, he returns to test P=1. He does this backtracking without any error or hesitation, and it is reasonable to assume he has a position marked clearly to which he intends to return. Without testing further values for P, he sets I=4, eliminating all values except 2 for P (presumably on the inference that P is less than I), and sets R at either 3 or 8. After testing 3, which proves unsuccessful, he tests 8, again without any hesitation, and solves the problem. This subject is distinguished by having a clear solution model 'in his head', which he searches, and which enables him to abandon an unsuccessful path without hesitation. His awareness of the rules of arithmetic are used directly to eliminate several values as being inappropriate for further testing. This protocol is very close to the 'ideal' protocol for this problem set out in Chapter 7.

PROTOCOL 5

This subject's performance was slightly below that of the mean solution time for Problem 2. He is similar to the subject of Protocol 4 in the relatively short time he spends in the first processing segment, and his second minute on the problem is marked by a high verbalization rate of about 2 words per second. Although he realizes that the setting of values for I plays a crucial role, his behaviour is much more haphazard than the subject above: he sets I equal to 7, for arbitrary reasons, he does not follow through all its implications, which, if he had, would have

provided him with a possible solution. He then tests I=9, then I=8, abandoning them without exhaustive testing or logical reason. He is clearly not convinced of the value of his method since he continues to search for other heuristic devices to apply, such as assigning a value to the product, looking for even and odd numbers etc. After this, he attempts to impose limits on the range of carries in the problem, and then concludes with the fact that I can equal 3,4,5,6 or 7. He then proceeds to process I=3, but does not focus on the letter I itself, looking rather at the implications this would have for the products, or for sub-units of the product. He avoids a straightforward application of the subroutine for varying the value of I, and searches for ways to impose limits on vectors of values. Still maintaining his hypotheses about divisibility and evenness-oddness, he sets I equal to 4, which eventually leads him to the solution. He is distinguished from the subject of Protocol 4 not so much by errors as by circumvention, searching for implications which are neither relevant nor useful.

Problem 3 (HANNAH)

PROTOCOL 6

This subject solved the problem in considerably less than the average time of 30.62 minutes for the Adult/Verbal group on this problem. He is similar to the subject of Protocol 7 both in solution time and output rate (about 60 words per minute). The protocol gives an adequate programme of his actions in that all major processing points appear, along with reasons for generating them. The regular processing rate suggests that this subject has an internal representation of a map in which his subgoals are laid out in orderly sequence. This may be a function of

his previous experience with the type of problem used. He ennumerates possibilities for carries, assuming, wrongly, that they may be equal to 2. He then manipulates even-odd relationships for D and N, then decides to "have a stab" setting N equal to 3, then to 2, and 4, and finally, correctly, to 5. He then decodes S, M and D. His short time in Module 3 is partly fortuitous, since he produced the right combination of digits for A, E and Y in the first recursion.

PROTOCOL 7

This subject is similar to the one of Protocol 6 not only in time and output, but also in his sporadic use of the display routine. He is distinguished by his concern to seek points of maximum information, and "symmetry" by which he is presumably referring to equations or pseudo-equations. He has gathered all the information he will need for processing the letters of Module 3 before he has assigned the letters of Module 2, and this manifestation of 'solution acceleration' presumably explains the short time he requires for processing Module 3, i.e. about one minute. There is a conspicuous absence of reference to the fact that three digits and three letters remain to be assigned (see Protocol 6) but the fact that he appears to know them suggests he has an externally represented blackboard and display to which he has access, and which he does not mention perhaps because the process of verbalization does not operate in that problem space.

PROTOCOL 8

This female student subject required considerably more time than the mean for her group of 42.18 minutes, and in fact was close to the cut-off point of sixty minutes that was used in the experiment. Her main difficulty arose from

being unable to employ the important heuristic method of restricting the domain of possible digits for a letter. In consequence, she tries any number haphazardly for any letter, but is not certain of those she has assigned. By not using either the elimination or the blackboard devices (Chapter 7), she is unable to make progress. Early on she says "Y can be any value\" but fails to realize that this characteristic entails the letter must be solved at the end of the problem. Consequently, she fixes on a value for Y, and then does not assign that value to any other letter. In other words, she develops a dysfunctional rigidity over the letter Y. The same rigidity appears in her approach to the letter U to which she assigns every letter except the correct one, zero, almost as if she does not consider zero to be a valid substitute. She appears to have no clear concept of the structure of the problem, or which are appropriate rules to apply in order to solve it. She accumulates more information than she can store or handle, and this is partly the cause of her inability to reject potential values for a letter on any rationale other than if it contradicts the value assigned to another letter. She does not consider any value to be inherently unsuitable because she has not delimited the potential domain for a letter by any logical or deductive means. Although she assigns the correct value of 4 to the letter Y while she is processing Module 2, she is not sure of it, and it is therefore accurate to describe the sequence in which the letters were correctly decoded as T-H-U-S-D-N-M-E-A-Y.

PROTOCOL 9

The same remarks are applicable to this protocol as were used to describe Protocol 2, except that the subject solved the problem in considerably less time than the mean time for the appropriate group.

REFERENCES

Anger, D.A. The dependence of interresponse times upon the relative reinforcement of different interresponse times. Journal of Experimental Psychology, 1956, 52, 145-161.

Backus, J.W. The syntax and semantics of the proposed international language of the Zurich ACM-GAMM Conference. Information Processing, UNESCO, 1959.

Bartlett, Sir F. Thinking: an experimental and social study. London: Allen and Unwin, 1958.

Bernard, J.R.L. Rates of utterance in Australian dialect groups. University of Sydney, Australian Language Research Centre, Occasional Paper No. 7, 1965.

Bharucha-Reid, A.T. Elements of the theory of Markov processes and their applications. New York: McGraw-Hill, 1960.

Binet, A. L'étude experimentale de l'intelligence. Paris: Schleicher Frères, 1903.

Bousfield, W.A. and Sedgwick, C.H.W. An analysis of sequences of restricted associative responses. Journal of General Psychology, 1944, 30, 149-165.

Bower, A.C. and King, W.L. The effects of number of irrelevant stimulus dimensions, verbalization, and sex on learning biconditional classification rules. Psychonomic Science, 1967, 8, 453-454.

Bower, G.H. Application of a model to paired-associate learning. Psychometrika, 1961, 26, 255-280.

Bower, G.H. and Trabasso, T.R. Concept identification. In E.D. Neimark and W.K. Estes (Eds.) Stimulus sampling Theory. San Francisco: Holden-Day, 1967, 499-522.

Brunk, L, Collister, E.G., Swift, Carolyn and Stayton, S. A correlational study of two reasoning problems. Journal of Experimental Psychology, 1958, 55, 236-241.

Bush, R.R. Estimation and evaluation. In R.D. Luce, R.R. Bush and E. Galanter (Eds.) Handbook of mathematical psychology. Vol. I, 1963, 429-469.

Bush, R.R. and Mosteller, I. Stochastic models for learning. New York: Wiley, 1955.

Bush, R.R. and Mosteller, F. A comparison of eight models In R.R. Bush and W.K. Estes (Eds.) Studies in mathematical learning theory. Stamford: Stamford University Press, 1959, 293-307.

Buswell, G.T. and Kersh, B.Y. Patterns of thinking in Solving Problems. University of California Publications in Education, Vol. 12, No. 2. Berkeley: University of California Press, 1956.

Chapman, D.C. Estimating the parameters of a truncated gamma distribution. Annals of Mathematical Statistics, 1956, 27, 498-506.

Claparède, E. La psychologie de l'intelligence. Scientia, 1917, 22, 353-368.

Davis, G.A. Current studies of research and theory in human problem solving. Psychological Bulletin, 1966, 1, 36-54.

Cohen, A.C. Estimating parameters of Type III populations from truncated samples. Journal of American Statistical Association, 1950, 45, 411-423.

Dansereau, D.F. and Gregg, L.W. An information processing analysis of mental multiplication. Psychonomic Science, 1966, 6, 71-72.

Davis, J.H. Models for the classification of problems and the prediction of group problem-solving from individual results. (Doctoral dissertation. Michigan State University, 1961).

Davis, J.H. The solution of simple and compound word problem. Behavioural Science, 1964, 9, 359-370.

Davis, J.H., Carey, M.H., Foxman, P.N. and Tarr, D.B. Verbalization, Experimenter presence and problem-solving. Journal of Personality and Social Psychology, 1968, 8, 299-302.

Davis, J.H. and Restle, F. The analysis of problems and prediction of group problem solving. Journal of Abnormal and Social Psychology, 1963, 66, 103-116.

Davis, M. Computability and unsolvability. New York: McCraw Hill, 1958.

Dennett, D.C. Machine traces and protocol statements. Behavioural Science, 1968, 13, 155-161.

Donaldson, Margaret. A study of children's thinking. London: Tavistock, 1963.

Duncan, C.P. Recent research on human problem solving. Psychological Bulletin, 1959, 56, 397-429.

Duncker, K. On problem solving. Psychological Monographs, 1945, 58(5), Whole No. 270.

Estes, W.K. Toward a statistical theory of learning. Psychological Review, 1950, 57, 94-107.

Feigenbaum, E.A. and Feldman, J. (Eds.) Computers and Thought. New York: McGraw Hill, 1963.

Feldman, J. Simulation of behaviour in the binary choice experiment. In E.A. Feigenbaum and J. Feldman (Eds.) Computers and thought, New York: McGraw Hill, 1963, 329-346.

Feller, W. An introduction to probability theory and its applications. Vol. I. New York: Wiley, 1960. (2nd Ed.)

Felsinger, J.M., Gladstone, A., Yamaguchi, H. and Hull, C.L. Reaction latency as a function of the number of reinforcements. Journal of Experimental Psychology. 1947, 37, 214-228.

Fisher, R.A. On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society, London, 1922, 222, 309-369.

Frijda, N.H. Problems of computer simulation. Behavioural Science, 1967, 12, 59-67.

Gagné, R.M. Problem solving and thinking. Annual Review of Psychology, 1959, 10, 147-172.

Gagne, R.M. and Smith, E.C. A study of the effects of verbalization on problem solving. Journal of Experimental Psychology, 1962, 63, 12-18.

Geternter, H. Realization of a geometry theorem-proving machine. In E. Feigenbaum and J. Feldman (Eds.) Computers and Thought. New York: McGraw Hill, 1963, 137-152.

Gelernter, H., Hansen, J.R. and Loveland, D.W. Empirical explorations of the geometry-theorem proving machine. In E.A. Feigenbaum and J. Feldman (Eds.) Computers and thought. New York: McGraw Hill, 1963, 153-163.

Green, B.F. Introduction: current trends in problem solving. In B. Kleinmuntz (Ed.) Problem solving: research, method and theory. New York: Wiley, 1966, 3-18.

Groot, A.D. de. Thought and choice in chess. The Hague: Mouton and Company, 1965.

Gur, J. An investigation into,and speculations about, the formal nature of a problem-solving process. Behavioural Science, 1960, 5, 39-59.

Hafner, A.J. Influence of verbalization in problem-solving. Psychological Reports, 1957, 3, 360.

Hayes, J.R. Memory, goals and problem-solving. In B. Kleinmuntz, (Ed.) Problem solving: research, method and theory. New York: Wiley, 1960, 149-170.

Hayes, W.L. Statistics for psychologists. New York: Holt, Rinehart and Winston. 1963.

Hull, C.L. Principles of behaviour. New York: Appleton-Century-Crofts, 1943.

Humphrey, G. Thinking: an introduction to its experimental psychology. New York: Wiley, 1963.

Hunt, C.C. and Kuno, M. Background discharge and evoked response of spinal inter-neurones.Journal of Physiology, 1959, 147, 364-384.

Hunter, I.M.L. The solving of five-letter anagram problems. British Journal of Psychology. 1959, 50, 193-206.

John, E.R. Contributions to the study of the problem-solving process. Psychological Monographs, 1957, 71, (Whole. No. 447).

Johnson, D.M. The psychology of thought and judgement. New York: Harper, 1955.

Johnson, E.S. An information-processing model of one kind of problem-solving. Psychological Monographs. 1964, (Whole No. 581)

Kaplan, I.A., Carvellas, T. and Breinin, Barbara J. Distribution of anagram solution times. Journal of Psychology, 1966, 62.

Kendall, M.G. and Stewart, A. The advanced theory of statistics. Vol. 1. London: C. Griffin, 1958.

Kilpatrick, J. Analyzing the solution of word problems in mathematics: an exploratory study. (Doctoral dissertation, Stamford University). Ann Arbor: Michigan University Microfilms, 1967. No. 68-6442.

Kleinmuntz, B. (Ed.) Problem solving: research, method and theory. New York: Wiley, 1966.

Köhler, W. The mentality of apes. New York: Harcourt Brace, 1925.

Lazerte, M.E. The development of problem solving in arithmetic. Toronto: Clarke, Irwin, 1933.

Lorge, I., Fox, D., Davitz, J. and Brenner, M. A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. Psychological Bulletin, 1958, 55, 337-372.

Luchins, A.S. Mechanization in problem solving. The effect of Einstellung. Psychological Monographs. 1942, 54, (Whole No. 248).

Maccoby, Eleanor E. Sex differences in intellectual functioning. In Eleanor E. Maccoby (Ed.) The development of sex differences. Stanford: Standord University Press, 1966, 25-55.

McCarthy, J. The inversion of functions defined by Turing machines. In C.E. Shannon and J. McCarthy (Eds.) Automa Studies, Annals of Mathematical Studies, Vol 34. Princeton, 1956.

McGill, W.J. Stochastic latency mechanisms. In R.D. Luce; R.R. Bush and E. Galanter (Eds.) Handbook of mathematical psychology, Vol. 1. New York: Wiley, 1963, 309-360.

Maier, R.F. Reasoning in humans: I. On direction. Journal of comparative psychology. 1930, 10, 115-143.

Means, Gladys H. The influence on problem-solving of following an efficient model. (Doctoral dissertation, University of Alabama). Ann Arbor, Michigan: University Microfilms, 1967, No. 67-1294.

Merton, R.K. Social theory and social structure. Rev. and enl. edition. New York: The Free Press, 1966.

Miller, G. Finite Markov processes in psychology. Psychometrika, 1952, 17, 149-167.

Miller, G.A., Galanter, F. and Pribram, K.H. Plans and the structure of behaviour. New York: Holt, 1960.

Minsky, M. Steps toward artificial intelligence. In E.A. Feigenbaum and J. Feldman (Eds.) Computers and thought. New York: McGraw Hill, 1963, 406-450.

Mood, A.M. Introduction to the theory of statistics. New York: McGraw Hill, 1950.

Moore, O.K. and Anderson, Scarvia B. Modern logic and tasks for experiments on problem solving behaviour. Journal of Psychology, 38, 151-160.

Moran, P.A.P. The statistical treatment of flood flows. Transactions of the American Geophysical Union, 1957, 38, 519-523.

Mueller, C.G. Theoretical relationships among some measures of conditioning. Proceedings of the Natural Academy of Sciences. 1950, 36, 123-130.

Nakamura, C.Y. The relation between conformity and problem-solving. Stanford: Department of Psychology, Stanford University, Technical Report No. 11. 1955. (Contract N6onr 25125)

Nance, R.D. and Sinnot, W. Learning factor in anagram solving. Psychological Reports. 1963, 13, 867-870.

Naylor, T.H., Balintty, J.L., Burdick, D.S. and Chujk. Computer simulation techniques. New York: Wiley, 1965.

Neisser, U. The imitation of man by machine. Science, 1963, 139, 193-197.

Newell, A. Some problems of basic organization in problem-solving programs. In M.C. Yovits, G.T. Jacobi and G.D. Goldstein (Eds.) Self-organizing systems. 1962. Washington: Spartan Books, 1962.

Newell, A. On the analysis of human problem-solving protocols. Carnegie Institute of Technology, 1966(a).

Newell, A. Discussion of papers by Robert M. Gagné and John R. Hayes. In B. Kleinmuntz (Ed.) Problem-solving: research, method and theory. New York: Wiley, 1966(b) 171-182.

Newell, A. Studies in problem-solving: Subject 3 on the crypt-arithmetic task. Donald and Gerald=Robert, Carnegie Institute of Technology, 1967.

Newell, A. and Ernst, G.W. The search for generality, In E.W. Kalenich (Ed) Proceedings IFIP Congress 65, Spartan, 1965. 17-24.

Newell, A. Shaw, J.C. and Simon, H.A. Elements of a theory of human problem-solving. Psychological Review, 1958(a), 65, 151-166.

Newell, A., Shaw, J.C. and Simon, H.A. The processes of creative thinking. Paper P-1320. Santa Monica: Rand Corp. 1958(b).

Newell, A., Shaw, J.C. and Simon, H.A. Report on a general problem-solving program. In Proceedings of the International Conference on Information Processing. Paris: UNESCO House, 1960, 256-264.

Newell, A. and Simon, H.A. Heuristic programs and algorithms. C.I.P. working paper 39. Carnegie Institute of Technology, 1962.

Paige, J.M. and Simon, H.A. Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.) Problem solving: research, method and theory. New York: Wiley 1966, 51-119.

Palzere, D.E. An analysis of the effects of verbalization in the learning of mathematics (Doctoral dissertation University of Connecticut). Ann Arbor, Michigan: University Microfilms, 1967. No. 68-1388.

Patrick, C. Creative thought in artists. Journal of Psychology. 1934, 4, 35-73.

Pearson, K. (Ed.) Tables of the incomplete Γ function. London: His Majesty's Stationery Office, 1922.

Phelan, J.G. A replication of a study on the effects of attempts to verbalize on the process of concept attainment. Journal of Psychology, 1965, 59, 289-293.

Poincaré, H. Science and method, (trans.F. Maitland). New York: Dover, 1952.

Polya, G. Mathematics and plausible reasoning. Princeton: Princeton University Press, 1954, 2 Vols.

Polya, G. How to solve it. (2nd Ed.) New York: Doubleday, 1957.

Polya, G. Mathematical discovery. New York: Wiley, 1962-65. 2 Vols.

Ray, W.A. Verbal compared with manipulative solution of an apparatus problem. American Journal of Psychology, 1957, 70, 289-290.

Reitman, W.R. Heuristic programs, computer simulation and higher mental processes. Behavioural Science, 1959, 4, 330-335.

Reitman, W.R. Information-processing models in psychology. Science, 1964, 144, 1192-1198.

Reitman, W.R. Cognition and thought: an information-processing approach. New York: Wiley, 1966.

Restle, F. The Psychology of thought and judgement. New York: Wiley, 1961.

Restle, F. Conditioning and discrimination: the all-or-none processes in paired-associate learning. Department of Psychology, Indiana University, technical report, Contract Nonr 908 (16) Bloomington, Indiana, 1962.

Restle, F. and Davis, J.H. Success and speed of problem solving by individuals and groups. Psychological Review. 1962, 69, 520-536.

Rimoldi, H.J.A., Fogliatto, Hermelinda M., Erdmann, J.B. and Donnelly, M.B. Problem-solving in high school and college students. Chicago: Loyola University Psychometric Laboratory, 1964 (Publication No. 42).

Roth, B. The effects of overt verbalization on problem-solving. (Doctoral dissertation, New York University). Ann Arbor, Michigan: University Microfilms, 1965. No. 65-9321.

Ruger, H.A. The psychology of efficiency. Archives de Psychologie, 1910, No. 15.

Samuel, A.L. Some studies in machine learning using the game of checkers. I.B.M. Journal of Research and Development, 1959, 3, 210-229.

Sargent, S.S. Thinking processes at various levels of difficulty. Archives de Psychologie, 1940, No. 249.

Siegel, S. Non-parametric statistics. New York: McGraw Hill, 1956.

Simon, H.A. and Barenfield, M. Information-processing analysis of perceptual processes in problem-solving. Psychological Review, 1969, 76, 473-484.

Skinner, B.F. An operant analysis of problem-solving. In B. Kleinmuntz (Ed.) Problem-solving: research, method and theory. New York: Wiley, 1966, 225-257.

Staats, A.W. An integrated-functional learning approach to complex human behaviour. In B. Kleinmuntz (Ed.) Problem-solving: research, method and theory, New York: Wiley, 1966, 259-339.

Stephan, F.F., and Mishler, E.G. The distribution of participation in small groups: an exponential approximation. American Sociological Review, 1952, 17, 598-608.

Sternberg, S. Retrieval of contextual information from memory. Psychonomic Science, 1967, 8, 55-56.

Suppes, P. and Ginsberg, Rose. A fundamental property of all-or-none models, binimial distribution of responses prior to conditioning with application to concept formation in children. Psychological Review, 1963, 70, 139-161.

Sweeney, E.J. Sex differences in problem-solving. Stanford: Department of Psychology, Stanford University, Technical Report No. 1, 1953. (Contract N6onr 25125)

Tocher, K.D. The application of automatic computers to sampling experiments. Journal of the Royal Statistical Society, 1954, B16, 39-61.

Weinberg, G.M. Experiments in problem-solving. (Doctoral dissertation. University of Michigan). University Microfilms Inc. Ann Arbor, Michigan, 1965. No. 66-6737.

Weinstock, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. Journal of Comparative and Physiological Psychology, 1954, 47, 318-322.

Wertheimer, M. Productive thinking. New York: Harper and Row, 1945.

Woodworth, R.S. and Schlosberg, H. Experimental psychology (revised). New York: Holt, 1954.